

Complex Networks Analysis and Visualization

Shtjefni Bryan

Contents

1 INTRODUCTION TO NETWORKS	11
1.1 Complex Systems	11
2 NETWORK STRUCTURAL ASPECTS	12
2.1 Zachary's Karate Club	12
2.2 Example of Retweet Networks	13
3 PROS AND CONS OF VISUALIZATION	13
4 BEHAVIOR AND DYNAMICS	13
5 INDIVIDUAL VS POPULATION LEVEL	15
6 INDIVIDUAL VS GROUPS	15
7 GRAPHS (Definitions)	16
7.1 Neighbors	17
7.2 Simple Graphs	17
7.3 Density and Sparsity	18
7.4 Subnetworks	18
7.5 Degree	18
7.6 Strength	19
7.7 Average Degree	19
7.8 Multilayer Networks	20
7.9 Temporal Networks	20
7.10 Networks of Networks	20
8 NETWORK REPRESENTATION	21
8.1 Adjacency Matrix	21
9 SPARSE NETWORK REPRESENTATION	22
9.1 Adjacency Lists	22
9.2 Edge Lists (with weights)	22
10 DRAWING NETWORKS	23
11 PATHS AND CONNECTIVITY	23
11.1 Paths	23
11.2 Shortest Paths	23
12 AVERAGE PATH LENGTH and DIAMETER	24
12.1 Average Path Length and Diameter with Disconnected Components	24
13 CONNECTEDNESS AND COMPONENTS	25
14 SMALL WORLD PHENOMENON	26
14.1 Milgram's Experiment	26
14.2 When a Path is really short?	26
15 A FRIEND OF A FRIEND - CLUSTERING COEFFICIENT	26
16 TRIADIC CLOSURE (Intro to Strength of Weak Ties)	27
16.1 Reasons For Triadic Closure	28
17 THE STRENGTH OF WEAK TIES	28
17.1 Bridges	28
17.2 Local Bridges	30
17.3 Different Level of Strength	31
17.4 Strong Triadic Closure	32
17.5 Local Bridges and Weak Ties - Proof	32

18 STRENGTH TIES AND LARGE SCALE DATA NETWORK	33
18.1 Contrast between Nodes	34
18.2 Misbehavior	34
18.3 Structural Hole - Social Capital	35
19 HOMOPHILY	35
19.1 Definition of Homophily	35
19.2 Birds of a Feather Flock Together	36
19.3 Measuring Homophily	36
20 DEGREE ASSORTATIVITY	39
20.1 Measuring Assortativity	40
20.2 Types of Assortativity	40
21 SELECTION AND SOCIAL INFLUENCE (Underlying Homophily)	41
21.1 Interplay between Selection and Social Influence	41
22 AFFILIATION	42
22.1 Affiliation Networks	42
22.2 Co-Evolution of Social and Affiliation Networks	42
23 CLOSURES	43
24 SHELLING MODEL	45
24.1 Spatial Model of Segregation	45
24.2 Formalization of Schelling Model	46
24.3 The Dynamics of Movement	46
24.4 Interpretations of the Model	46
25 NETWORK MEASURES	47
25.1 Real Networks are Heterogeneous	47
26 CENTRALITY MEASURES	47
26.1 Degree Centrality	47
26.2 Closeness Centrality	48
26.3 Betweenness Centrality	49
27 CENTRALITY DISTRIBUTIONS	50
27.1 Cumulative Distributions	51
27.2 Logarithmic Scale	51
27.3 Heterogeneity Parameter (Degree Distribution)	52
28 FRIENDSHIP PARADOX	53
28.1 Where does the Friendship Paradox come from?	53
29 ULTRA SMALL WORLD	54
30 ROBUSTNESS	54
30.1 How can we check Robustness?	54
31 CORE DECOMPOSITION	56
32 BETWEENNESS AND GRAPH PARTITIONING	57
32.1 Divisive Approach	57
32.2 Agglomerative Approach	57
32.3 Which edge to Remove First?	57
32.4 But if there are no Local Bridges?	57
33 GIRVAN NEWMAN ALGORITHM	58
33.1 Computing Edge Betweenness Efficiently	58
33.2 Consideration of Girvan Newman Algorithm	60
34 NETWORK MODELS	61

35 FEATURES OF REAL NETWORKS	62
35.1 Small World Property	62
35.2 High Clustering Coefficient	62
35.3 Heterogeneity	62
36 RANDOM NETWORKS	63
36.1 Erdos-Renyi Algorithm	63
36.2 Evolution of Random Networks	63
37 NUMBER OF LINKS, DENSITY AND AVERAGE DEGREE IN RANDOM NETWORKS	65
37.1 Expected Number of Links of a Random Network with N nodes	65
37.2 Expected Density of Links d of a Random Network with N nodes	65
37.3 Expected Average Degree $\langle k \rangle$ of a Random Network with N nodes	66
37.4 Degree Distribution in Random Networks	66
37.5 Small World Property in Random Networks	68
37.6 How many steps does it take to cover the whole network?	68
37.7 Clustering Coefficient in Random Networks	69
38 SUMMARY OF RANDOM NETWORKS	69
39 SMALL WORLD NETWORKS	70
39.1 The Watts-Strogatz Model	71
39.2 Degree Distribution of the Watts-Strogatz Model	72
39.3 Summary of the Watts-Strogatz Model	72
40 THE CONFIGURATION MODEL	73
41 PREFERENTIAL MODELS	74
41.1 Network Growth	74
41.2 Preferential Attachment	74
41.3 Polya's Urn Model	74
41.4 The Barabasi-Albert Model	75
41.5 Other Preferential Models	76
42 FINAL CONCLUSIONS ABOUT MODELS	76
43 COMMUNITIES IN NETWORKS	77
43.1 Community Structure	77
43.2 Why to Study Communities?	77
43.3 Definition of Variables	78
43.4 Definitions of Community	78
43.5 Community Definitions Based on Cohesion	79
43.6 Community Definitions Based on Cohesion vs Separation	79
43.7 Better Definition of Communities	80
44 PARTITIONS	81
44.1 Overlapping Communities	81
44.2 Hierarchical Communities	82
44.3 Network Partitioning	82
44.4 Related Problems in Network Partitioning	83
44.5 Network Bisection	83
45 Kernighan-Lin Algorithm	83
45.1 Limits of Network Partitioning	84
46 DATA CLUSTERING	85
47 HIERARCHICAL CLUSTERING	85
47.1 Structural Equivalence	85
47.2 Similarity of Nodes Groups	85
47.3 Types of Hierarchical Clustering	86
47.4 Dendograms	86
47.5 Limits of Hierarchical Clustering	87

48 COMMUNITY DETECTION	87
49 BRIDGE REMOVAL	87
49.1 Girvan Newman Algorithm for Bridge Removal	87
49.2 Quality Function (How to determine how good a partition is?)	88
50 MODULARITY	88
50.1 Homophily Test and Modularity	88
50.2 Formula of Modularity	89
50.3 Features of Modularity	90
50.4 Modularity Optimization	90
50.5 Newman's Greedy Algorithm	90
50.6 Louvain Algorithm	91
50.7 Considerations on the Louvain Algorithm	92
50.8 Limits of Modularity Optimization	92
50.9 Resolution Limit	93
51 LABEL PROPAGATION	94
51.1 Limits of Label Propagation	95
51.2 Good Things about Label Propagation	95
52 STOCHASTIC BLOCK MODELING	95
52.1 Fitting a Stochastic Block Model to a Network	97
52.2 Limits of Stochastic Block Modeling	98
53 METHOD EVALUATION	98
53.1 Planted Partition Model	99
53.2 Girvin Newman - GN Benchmark	100
53.3 LFT Benchmark	100
53.4 Negative Tests	101
53.5 Real Benchmark	101
53.6 Partition Similarity	102
53.7 Fraction of Correctly Detected Nodes	102
53.8 Normalized Mutual Information	102
54 SIGNED NETWORKS	103
54.1 Segregation and Polarization	103
54.2 Issues with Studying Polarization	105
55 STRUCTURAL BALANCE	106
55.1 Structural Balance Properties	107
55.2 The Balance Theorem	107
55.3 Proof of The Balance Theorem	108
55.4 Applications of Structural Balance	109
55.5 A Weaker Form of Balance	110
55.6 Weak Structural Balance Property	110
55.7 Weakly Balance Theorem and its Proof	111
55.8 Generalizing the Definition of Structural Balance	112
55.9 Claim on Arbitrary Non Complete Graphs	113
55.10BFS of the Reduced Graph	114
55.11Recap of the Proof (Negative Cycle in BFS)	114
56 GAME THEORY	115
56.1 What is a Game?	115
56.2 Students Example	115
56.3 Basic Ingredients of a Game	116
56.4 Reasoning about Behavior in a Game	116
56.5 Prisoner's Dilemma	117
56.6 Formalization of the Best Responses	118
56.7 Formalization of Dominant Strategies	118
56.8 What if only one Player has a Strictly Dominant Strategy?	119
56.9 What if no one has a (Strict) Dominant Strategy?	119

56.10	Nash Equilibrium	119
56.11	Multiple Equilibrium	119
56.12	What if Two Players don't Agree	120
56.13	The Stag-Hunt Game	120
56.14	No Nash Equilibrium	120
56.15	Mixed Strategies	121
56.16	Payoffs for Mixed Strategies	121
56.17	Equilibrium with Mixed Strategies	121
56.18	Interpretation of The Indifference Principle	121
56.19	Optimalities – Pareto Optimality – Social Optimality	122
56.20	Networks and Game Theory	122
57	TRAFFIC AT EQUILIBRIUM	123
57.1	Traffic Game	123
57.2	Equilibrium Traffic	123
57.3	Can we have a Nash Equilibrium without an even balance?	124
58	BRAESS'S PARADOX	124
58.1	The Social Cost of Traffic at Equilibrium	124
58.2	Traffic Pattern	125
58.3	Finding a Traffic Pattern at Equilibrium	125
58.4	Does a best-response dynamics always stop?	125
58.5	Progress Measure	125
58.6	Potential Energy	126
58.7	Does the best-response dynamics stop?	126
58.8	Comparing Equilibrium traffic to the Social Optimum	127
58.9	Potential Energy and Total Travel Time (TTT) for an Edge	128
58.10	Total Travel Time at Equilibrium and at Social Optimality	129
58.11	Conclusions of Traffic Game	129
59	DIRECTED NETWORKS	130
59.1	Examples of Directed Networks	131
60	THE WEB AND ITS STRUCTURE	131
60.1	The Web as a Directed Graph	131
60.2	The Structure of The Web	132
60.3	The Bow Tie Structure of The Web	133
60.4	Web Crawlers	133
60.5	Which Page is a Good Seed Page?	135
60.6	Topical Locality	136
60.7	Topic Locality and Topic Drift	137
60.8	Cosine Similarity	138
60.9	Pearson's Similarity	140
61	WEIGHTED NETWORKS	141
61.1	Information Diffusion Networks	141
61.2	Retweet Networks	141
61.3	Virality	142
61.4	Influence	142
61.5	Co-occurrence Networks	143
61.6	Weight Heterogeneity	143
61.7	Link Filtering	144
62	SEARCHING THE WEB	145
62.1	The Problem behind Searching The Web	145
63	HITS – LINK ANALYSIS USING HUBS AND AUTHORITIES	146
63.1	Finding Lists and List Value	146
63.2	Calculating Hubs and Authorities	146

64 SPECTRAL ANALYSIS OF HITS	148
64.1 Update Rules	149
64.2 Understanding the k-step Hub Authority Computation	150
64.3 Introducing Normalization – Multiplications and Eigenvectors	151
64.4 Eigenvectors and Square Matrices	152
64.5 Convergence of the Hub Authority Computation	153
64.6 Final Considerations	154
65 PAGE RANK	155
65.1 Page Rank and Stabilization	155
66 SPECTRAL ANALYSIS OF PAGE RANK	156
66.1 Using Matrix Notation	156
66.2 Update Rule	157
66.3 Repeated Improvement	157
66.4 Convergence of the Scaled PR Update Rule	158
67 RANDOM WALKS AND PAGE RANK	158
67.1 Leakage	159
67.2 Formulation of The Page Rank using Random Walks	159
67.3 A scaled Version of the Random Walk	160
68 INFORMATION CASCADES - FOLLOWING THE CROWD	160
68.1 The Milgram Experiment on Obedience	160
68.2 Collective Behavior	161
68.3 The Milgram Experiment on Conformity and Independence	161
68.4 Information Cascade	161
69 A SIMPLE HERDING EXPERIMENT	161
69.1 Asch Conformity Experiment	161
69.2 Herding Experiment (Herding, urns and marbles)	162
70 BAYES RULE – A model of Decision Making under Uncertainty	163
70.1 Bayes Rule in the Herding Experiment	165
71 A SIMPLE CASCADING MODEL	167
71.1 Model Goals	167
71.2 Model Ingredients and Assumptions	167
71.3 Final Considerations of the Herding Experiment	168
71.4 Individual Decisions	168
71.5 What about the Payoff?	169
71.6 Multiple Signals	170
72 SEQUENTIAL DECISION MAKING AND CASCADES	172
72.1 What Happens when the Number of Individuals Grows to Infinity?	173
73 LESSONS FROM CASCADE EFFECTS	173
74 NETWORK EFFECTS	174
74.1 Externalities	174
75 THE ECONOMY WITHOUT NETWORK EFFECTS	174
75.1 The Market for a Good	174
75.2 Reservation Prices	175
75.3 Market Price	175
75.4 The Equilibrium Quantity of the Good	176
76 THE ECONOMY WITH NETWORK EFFECTS	177
76.1 Perfect Predictions Hypothesis	177
77 STABILITY, INSTABILITY AND TIPPING POINTS	178

78 A DYNAMIC VIEW OF THE MARKET	180
78.1 Modeling the Population Reaction	180
78.2 Analyzing the Dynamics	181
79 INDUSTRIES WITH NETWORK GOODS	182
79.1 Lowering The Production Cost	182
79.2 Marketing a Product with Network Effects	182
79.3 Social Optimality with Network Effects	182
79.4 Competition	182
80 MIXING INDIVIDUAL AND POPULATION LEVEL EFFECTS	183
80.1 Growing an Audience from Zero	183
80.2 Bottlenecks and Large Changes	184
81 CONSEQUENCES OF SMALL CHANGES	184
82 POWER LAWS AND RICH GET RICHER PHENOMENA	185
83 POPULARITY AS A NETWORK PHENOMENON	185
83.1 Popularity, Heterogeneity and Networks (Recap)	185
83.2 Case Study: The Web	185
83.3 Looking For a Popularity Scale	186
84 POWER LAWS	187
84.1 Empirical Findings for The Web – Power Law	187
84.2 The Lack of a Scale	188
84.3 Are Scale Free Networks Rare or Frequent?	188
84.4 Power Laws	189
84.5 Fitting with a Straight Line	189
84.6 Why Do Hubs Emerge?	189
85 RICH GET RICHER MODELS	190
86 THE UNPREDICTABILITY OF THE RICH GET RICHER EFFECTS – IMPACTS OF POWER LAW	190
86.1 The Fragility of Popularity	190
86.2 The Music Lab Experiment	191
87 THE LONG TAIL	192
87.1 Zipf's Law	193
87.2 Pareto's Law	193
88 THE EFFECT OF SEARCH TOOLS AND RECOMMENDATION SYSTEMS	194
89 ANALYSIS OF THE RICH GET RICHER MODEL	194
89.1 Objectives	195
89.2 Recap of The Rich Get Richer Process	195
89.3 The Deterministic Argument	196
89.4 Identifying a Power Law in the Deterministic Approximation	198
89.5 Final Step	199
89.6 Meaning of the 'c' Exponent in function of p	200
89.7 Conclusion	200
90 CASCADE BEHAVIORS IN NETWORKS	201
91 DIFFUSION IN NETWORKS	202
91.1 Individual's Choices Depending on Others	202
91.2 What if we Consider Heterogeneity?	202
91.3 The Diffusion of Innovations	203

92 MODELING DIFFUSION THROUGH A NETWORK	203
92.1 Diffusion of a New Behavior	203
92.2 A Network Coordination Game	204
92.3 Cascading Behaviors	204
92.4 Monotonically Spreading	205
93 CASCADES AND CLUSTERS	206
93.1 Stopping Cascades	206
93.2 Relationship Between Clusters and Cascades	207
93.3 Clusters Imply No Cascade - No Cascade Imply Clusters	207
94 DIFFUSION, THRESHOLD AND THE ROLE OF WEAK TIES	208
94.1 Viral Marketing	208
94.2 The Role of Weak Ties	208
94.3 Simple vs Complex Contagion	209
94.4 Centola's Experiment	209
94.5 Finding The Optimal Clustering For Spreading	210
95 EXTENSION OF THE BASIC CASCADE MODEL	211
96 KNOWLEDGE, THRESHOLDS AND COLLECTIVE ACTION	211
96.1 Collective Actions and Pluralistic Ignorance	212
96.2 A Model for The Effect of Knowledge on Collective Actions	212
96.3 Common Knowledge and Social Institutions	212
97 THE CASCADE CAPACITY	213
97.1 Infinite Networks	213
97.2 How Large Can the Cascade Capacity Be?	213
97.3 The Size of the Interface	214
97.4 Another Extension of the Model	214
98 SMALL WORLD AND DECENTRALIZED SEARCH	214
98.1 Recap of Milgram's Experiment and Six Degree of Separation	214
98.2 Structure And Randomness	214
98.3 Evolution of Random Networks	215
98.4 How many steps does it take to cover the whole network?	216
98.5 Watts-Strogatz Model Recap and Alternative Formulation	217
99 DECENTRALIZED SEARCH	219
99.1 Searching for Short Paths in a Decentralize Way	219
99.2 Distances From The Target are Constantly Reduced	219
99.3 An Algorithmic Perspective (GeoGreedy Algorithm)	220
99.4 Navigation in a Small World	220
99.5 Comparing Models with Different Clustering Exponents	221
99.6 Estimate of the Clustering Exponent q by Simulation	221
99.7 Intuitive Explanation of $q = 2$	222
99.8 Qualitative Interpretation of $q = 2$	223
100 EMPIRICAL ANALYSIS AND GENERALIZED MODELS	223
100.1 Rank-Based Friendships	223
100.2 Ranks and GeoGreedy	224
100.3 Social Foci and Social Distance	224
101 CORE PERIPHERY STRUCTURES AND DIFFICULTIES IN DECENTRALIZED SEARCH	225
102 ANALYSIS OF DECENTRALIZED SEARCH	226
102.1 Time Spent in One Phase	229
102.2 Analysis in Two Dimensions	232
103 EPIDEMICS - DISEASES AND THE NETWORKS THAT TRANSMIT THEM	233
103.1 Contact Networks	233
103.2 Simple vs Complex Contagion Models	233

104 BRANCHING PROCESS	234
104.1 The Basic Reproductive Number r_0	235
105 THE SIR EPIDEMIC MODEL	236
105.1 SIR Infections in Networks	237
105.2 The Role of r_0	238
105.3 SIR Epidemics and Percolation	238
106 THE SIS EPIDEMIC MODEL	239
106.1 Connections Between SIR and SIS	240
107 SYNCHRONIZATION AND OSCILLATIONS	241
108 THE SIRS EPIDEMIC MODEL	241
108.1 Small World Contact Networks	242
109 ANALYSIS OF THE BRANCHING PROCESS	243
109.1 The Claim we want to Prove	243
109.2 Following the Values Q_n to a Limit	245
110 EPIDEMICS SPREADING IN HETEROGENEOUS NETWORKS	248
110.1 Real Networks	248
111A CONTINUUM APPROACH TO MODEL EPIDEMICS	250
111.1 Revisiting Compartmental Modeling	250
111.2 SI Model	250
111.3 Logistic Growth Curve	252
111.4 Observation of SI Epidemics	252
111.5 SIS Model	252
112 SIR Model	254
112.1 Calculating the Asymptotic Value of Recovered Individuals r	255
113 HETEROGENEOUS NETWORKS EPIDEMICS	255
113.1 Removing the Full Mixed Hypothesis	255
113.2 Degree-Block Approximation	255
113.3 SI Model on a Network	256
113.4 Observation on the Characteristic Time τ	257
113.5 SIS Model in a Network	258
113.6 Epidemic Threshold	258
113.7 No Epidemic Threshold and Hubs	260

1 INTRODUCTION TO NETWORKS

1.1 Complex Systems

A complex system doesn't mean a complicated system. A complex system is composed by many **interactive elements**. In complex systems we focus on the collective phenomena that emerges from the multitude of interactions present in the system. **Emergence** is not directly related to an individual phenomena but it manifests in the system with many interacting elements, even all of them.

Another important aspect is the difference between **linearity and non linearity**. In physics systems we have linear laws that describe systems, if we combine the laws with statistical elements, we see that these laws are linked with statistical distributions that characterize the laws themselves. With non linearity we don't have something like an average to describe the law we are working with, but we can have a positive or negative interval that is not linear, that can maybe follow an exponential law. When we move from linearity to non linearity things get complex. Non linearity is very common in real life.

Non linearity is also linked with **heterogeneity**. If we have a phenomena characterized by a Gaussian we have a so called **homogeneous behaviour**. When we have heterogeneity it means that we can have values that vary much from each other, like earthquake value etc.. they follow non linear distribution and have therefor an heterogeneous behaviour.

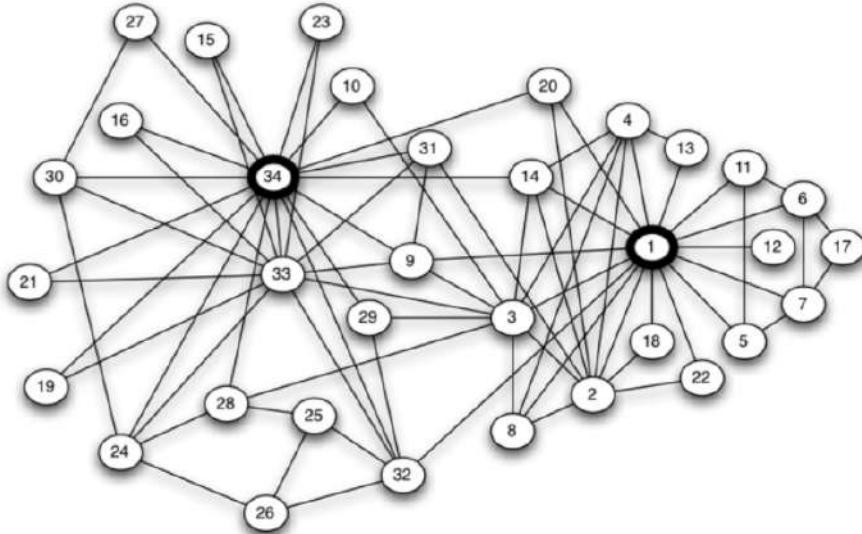
Interactive elements can change the behaviour of the other elements. Interactive elements can show something that cannot be predicted by individual behaviour, but the interaction itself produces the manifestation of something new. What we observed with the metronomes, we can show that we have interconnection of small units from a local level. Like a block of birds that flocks away together, they can do that because they communicate with each other, if they don't do that they aren't synchronized.

Networks are everywhere, we have social networks where actors (individuals, agents) interact with each other, we have social ties as well. We have information systems like books and web pages. Protein interaction networks, technical and economic systems.

2 NETWORK STRUCTURAL ASPECTS

We have a trivial representation of complex systems like stars, rings and grid topologies, where few characteristics describe the networks. In complex networks we don't have such simple characteristics to describe the phenomena, and we need a language or a framework to describe the complex Network

2.1 Zachary's Karate Club

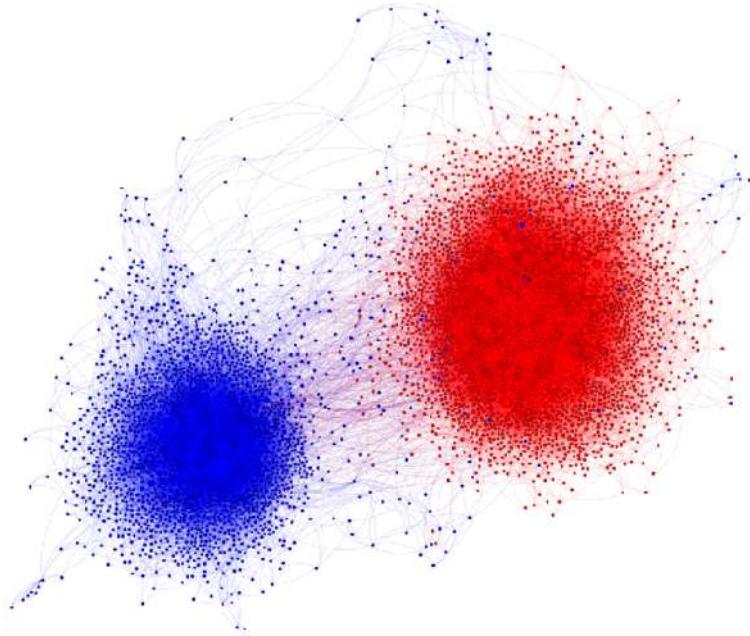


It connects many different people enrolled to the club and describes only the friendship relationships between the single individuals. This is not a star topology, this is a complex topology and is difficult to describe it.

We can still find some regularities, if we look at node 1 and 34, they are much more central compared to the other node, they have more connections than the others. This is a type of heterogeneity since we can assume that the node 1 and 34 are the dominant leaders in this social networks that are not liked with each other. This network looks kinda 'random', and we need to understand what is randomness in the networks.

If we take the internet as an example, it looks chaotic and random, but we can still see an apparent order emerging, since we can see the 'core vs periphery structure'. We need a language to describe the regularities.

2.2 Example of Retweet Networks



Red nodes are blogs that lead by republicans, while blues are democrats. If we look at the retweet system we see two different and polarized **echo chambers**, two different communities. This indicates that two people with the same political ideas are more likely to be connected. This phenomena is called **homophily** and is the tendency of individuals to link with similar ones, because if the people link with other people with the same idea as mine, it's likely that i wont change my mind.

3 PROS AND CONS OF VISUALIZATION

Visualizing a network suggests an inherent complexity, it says a lot about the structure we are observing but nothing about the reasons of why things are organized in that way. Is very difficult to summarize the whole network because there are differences and factors to take into consideration like core vs periphery structure, hierarchical structure, tightly-linked regions. We need a language to understand if we find some behavior, links, and such and give them a name.

4 BEHAVIOR AND DYNAMICS

After we understand how to name different characteristics in the network, we can define a connectedness in the graph and we can focus on the structure and the behavior of a node. Each node can have interests, reasons and tendencies to link to other nodes. This behavior must be observed in a time period because we want also to understand how such behaviors can change in time. Each individual's action have consequences for the outcomes of everyone in the system, if I'm a leader of a republican group, and change my beliefs and become a democrat, this will have consequences for the people that followed me. People can still follow me and change belief or not and unfollow.

We need a framework for reasoning about behavior and interactions in networks, we need something that takes into account a rational behavior that sometimes lead to strategies, and reasoning on strategies that may change. We will use some principals to evaluate individuals actions at a network level, when they interact with each other. Actions are not evaluated at individual level, but always on a network level.

In large populations we have the emergence of new ideas, beliefs, opinions, innovations, technologies, products, and social conventions but we need to determine when they establish as social practices. We live in a world where individuals can influence others or be influenced by others, but at the surface we know from psychology studies that humans have a tendency to conform to ideas or events.

We also know that individuals may have an interest in adopting the behavior of their neighbors, and this is called **cascading behavior**, and we can observe this in social contagion or in epidemics. **The social contagion** is based on the influence of the adoption of a new product or idea that can cascade throughout the whole network, and we can have some individuals that continue to spread the idea, called **spreaders**, or other closed communities that can stop the contagion.

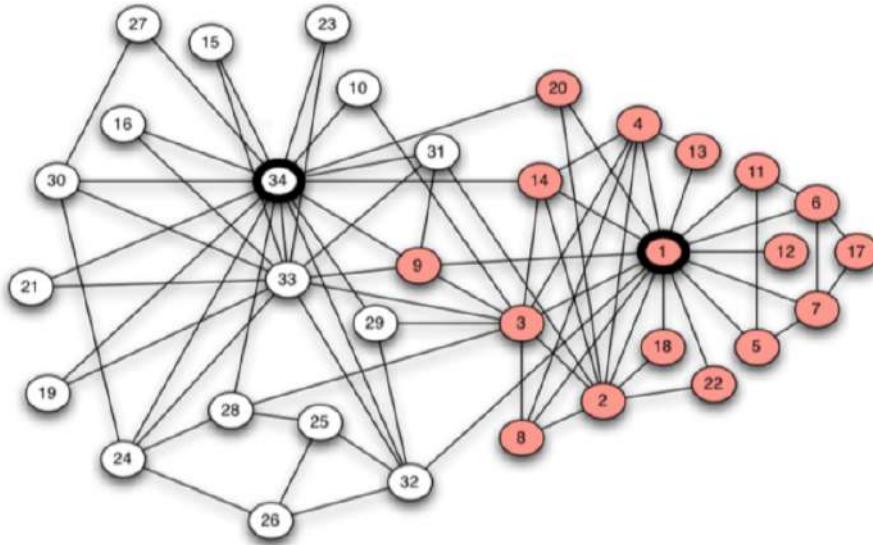
We can go from **raw data to networks** since we have access to massive amounts of network datasets. Data driven models allow us to make predictions but we cannot generalize this predictions since networks can be different.

5 INDIVIDUAL VS POPULATION LEVEL

When a large group of people is interconnected, they often respond in complex ways that can be observable only at the population level. **Viral Ideas** for example become viral because a large group of 'connected' people interact with it

At individual level sometimes an information goes viral or one person becomes very popular, some other times this does not happen, because we have no predictions at this level. An example is to measure the successfullness of a person, this is so very difficult to do.

6 INDIVIDUAL VS GROUPS



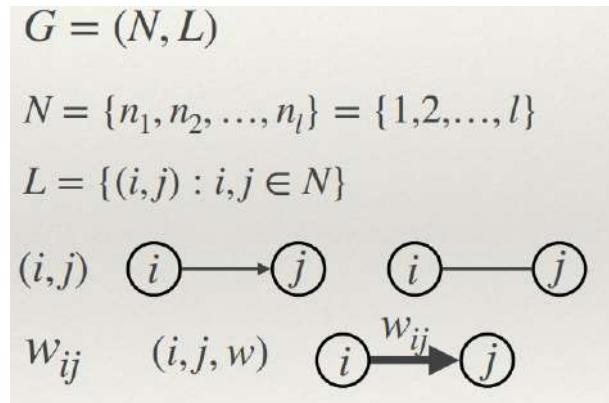
We can find clues to the division that eventually split the group into two separate clubs. We can have different roles. In this graph some people follow one club, and others other. We want to understand if this partition can be predicted algoritmically, we want kinda to predict the formation of communities.

7 GRAPHS (Definitions)

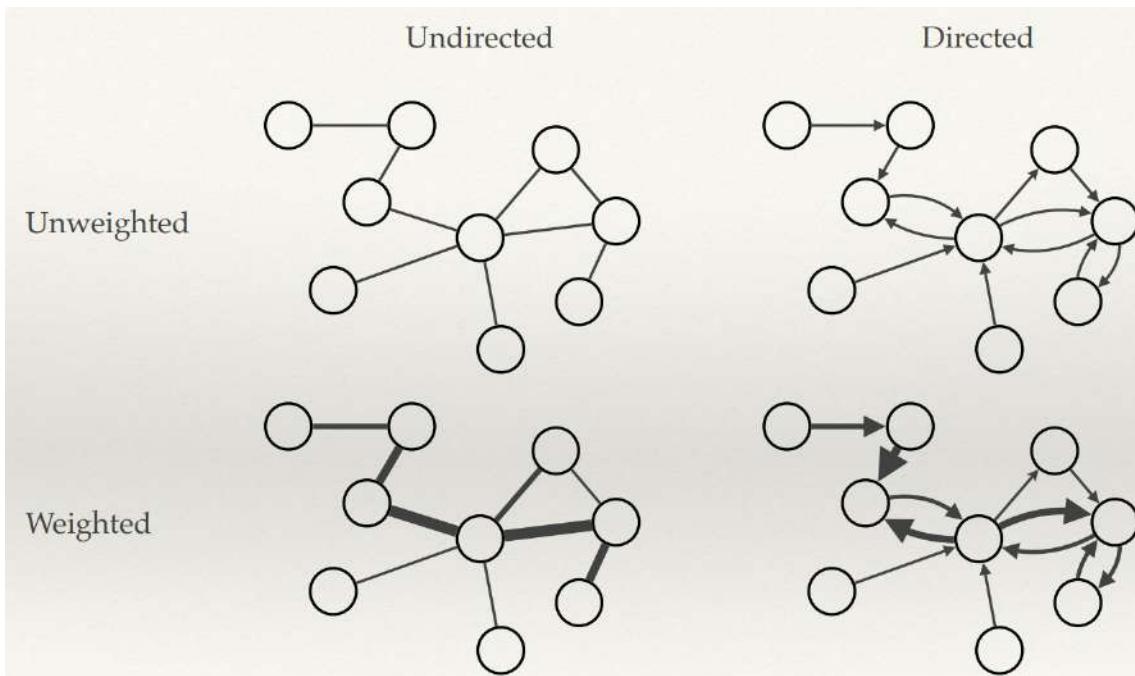
We use **graphs** because they are a **mathematical model** that helps to represent complex systems in terms of nodes and links. Whenever we have relationships between individuals that need to be represented somehow, we can use graphs because they are the most straight forward method of representation.

Graphs provide us with basic **languageses** to understand the basic elements of a network, and with this languages we can characterized precisely the **structure** and the **behavior** of networks. The networks can have structures that can have an **impact on the processes** that I can observe in the network.

A graph is a pair $G(N,L)$ of a set N of nodes and L of links. **Nodes** (vertexes) are just lists of objects, **Links** are pairs (i,j) , and the set L contains all the links that are formed inside the network.



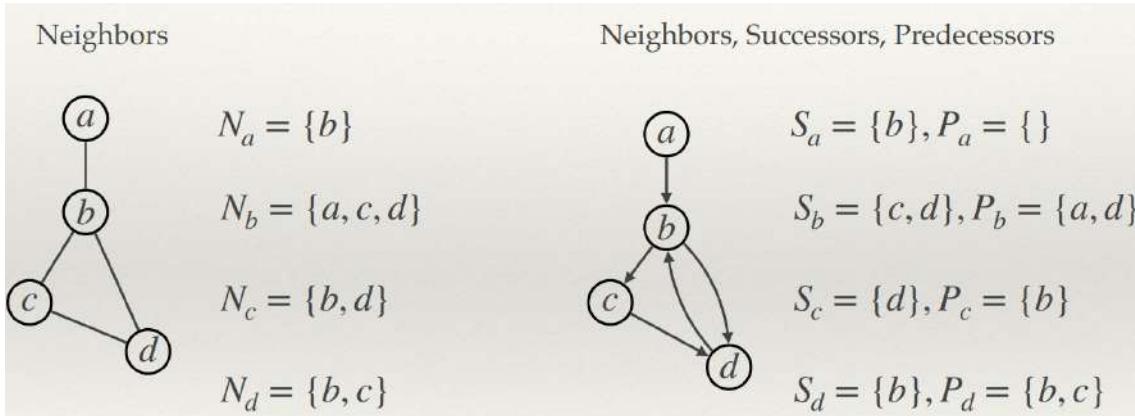
Graphs can be **directed or undirected**, in directed graphs we represent the link with arrows and in undirected is just a simple line. Graphs can also be **weighted or unweighted**, and we represent the weight with W_{ij} that is the weight of the link that connects i and j .



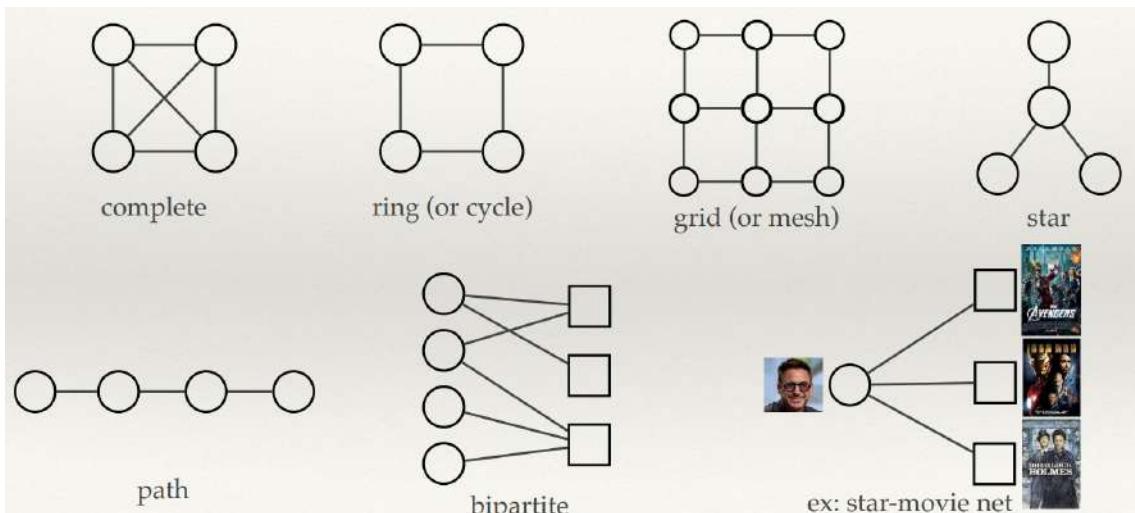
The thickness of a link states the weight of the link inside the network.

7.1 Neighbors

If we have a node inside a graph, we have a set of neighbors. Neighbors can be also **successors** or **predecessors**.



7.2 Simple Graphs



If every node of a network is connected to all the other nodes in the network, the set of neighbors of a single node is N – that given node, we have a **complete** graph. We can have other types of simple graphs such as rings, grids, stars, paths, bipartite etc.. Simple graphs are characterized by simple rules that allow us to define their characteristics.

7.3 Density and Sparsity

❖ Network size = number of nodes	$ N $ or, simply N
❖ Number of links	$ L $ or, simply L
❖ Maximum possible number of links	$L_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$
❖ Density $d = \frac{L}{L_{\max}} = \frac{2L}{N(N-1)}$	
❖ Sparsity if $d \ll 1 \Rightarrow$ sparse	$\begin{cases} L \text{ in the order of } N & \Rightarrow \text{sparse} \\ L \text{ in the order of } N^2 & \Rightarrow \text{dense} \end{cases}$

The **size** of a network is given simply by the number of nodes (N) inside the network. The number of links is instead defined with L , and with this we can define the **maximum number of links** (L_{\max}) that is given by the binomial coefficient (N chooses 2), and this formula obviously applies only to undirected graphs.

Density (d) is given by the fraction of L and L_{\max} . This is a simple measure that allows us to estimate the density of links in a network.

If d is much lower than 1, we say that the graph is **sparse**. To be more precise, if L is in the order of N then we have that the graph is sparse, and if it is in the order of N to the power of 2 it is **dense**.

7.4 Subnetworks

If we have a network, we can extract some nodes to create some subnetworks. A **subnetwork** is obtained by selecting a subset of the nodes and all the links among these nodes. A subnetwork that is also a complete network, is called a **clique**.

7.5 Degree

Number of links (or neighbors)	In directed networks
$i \rightarrow N_i \quad k_i = N_i $ degree	$k_i^{in} = P_i $ in-degree
Singleton: a node whose degree is zero	$k_i^{out} = S_i $ out-degree
$N_i = \{\}, k_i = 0$	$k_i = k_i^{in} + k_i^{out}$

The **degree** is given by the number of links (or neighbors) of a determined node i . A **singleton** is a node that has a degree equal to 0, and this means that this node is totally isolated from the rest of the graph. In directed networks we can have **in-degree**, that is the number of incoming links inside of a node, and **out-degree** that is the number of outgoing links of that node. If we sum the in and out degree we obtain the total degree of the node i .

7.6 Strength

Strength: Weighted degree

$$s_i = \sum_{j \in N_i} w_{ij}$$

in-strength

$$s_i^{in} = \sum_{j \in P_i} w_{ji}$$

out-strength

$$s_i^{out} = \sum_{j \in S_i} w_{ij}$$

If we sum not only the number of link, but also the weights of these links, we can calculate the **strength** of a determined node. The strength (s_i) is given by the sum of w_{ij} with j being a neighbor of i . The **in-strength** is the sum of all j 's connected to i and where j is a predecessor of i . The **out-strength** is the same but we have that j is a successor of i .

7.7 Average Degree

- ◊ average degree

$$\begin{aligned} <k> &= \frac{\sum_{i \in N} k_i}{N} \\ (\text{Undirected Net}) &= \frac{2L}{N} & d &= \frac{2L}{N(N-1)} \Rightarrow L = \frac{dN(N-1)}{2} \\ &= \frac{dN(N-1)}{N} & &= d(N-1) \end{aligned}$$

- ◊ Average degree is also connected to density

$$d = \frac{<k>}{N-1} = \frac{<k>}{k_{max}}$$

Is a measure of the full graph because we calculate the average of all the nodes of the network. The **average degree** $<k>$ is the sum of all the degrees of the nodes (k_i) divided by the size of the network N . If we sum all the degrees we count all the edges twice, so we can calculate the avg degree with $2L / N$. We can substitute L , using the formula of density, inside the equation and we can obtain that the average degree is exactly equal to the density multiplied by $N-1$ ($d(N-1)$). So we have that avg degree and density are connected with each other.

7.8 Multilayer Networks

A network can have **multiple layers** with its own nodes and edges (air transportation networks of distinct airlines, with some but not complete overlap of airport nodes). **Intralayer** links are links among nodes in the same layer, while **interlayer** link nodes across two different levels. Interlayer links are couplings linking the same node across layers (layers to represent different types of relationship in a social network, such as friendship, family ties, coworkers). If the sets of nodes in the different layers are identical, the network is a **multiplex**.

7.9 Temporal Networks

Temporal Networks is a multiplex in which the layers represent links at different times, called **temporal snapshots**. An example of this could be the retweets of a given tweet. This is an interesting way to study cascade effects inside of networks.

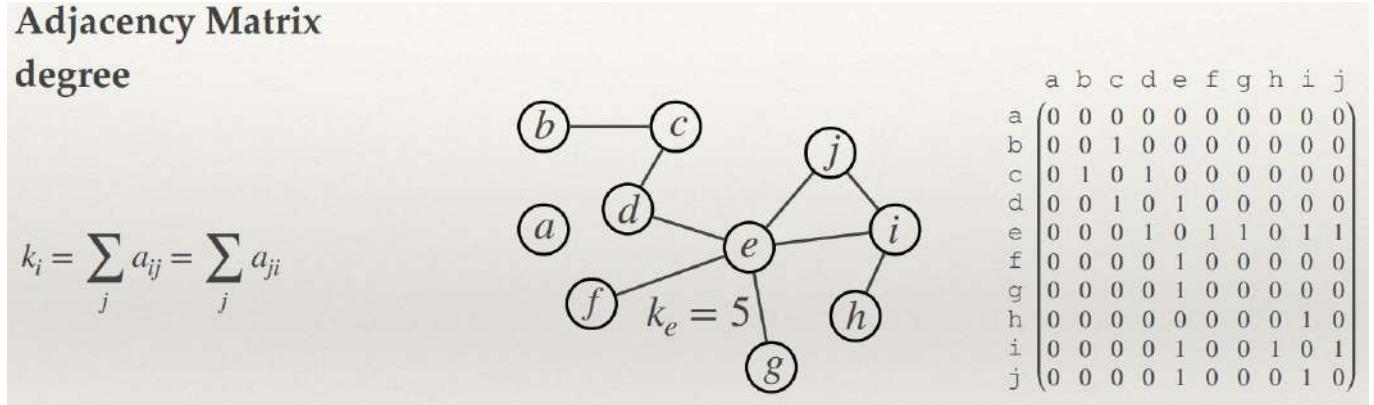
7.10 Networks of Networks

Each layer in a multilayer network can have its own nodes and edges, this is called a **network of networks** (the electrical power grid and the Internet). The interactions between networks in different layers are captured by interlayer links (power stations communicate via the Internet, Internet routers are powered by the power grid). Inside this scenario we can run into **cascading failures**, which are unpredictable vulnerabilities that manifest themselves in these networks of networks, since a problem of a determined network could cause massive problems to another network.

8 NETWORK REPRESENTATION

8.1 Adjacency Matrix

Adjacency Matrix is a matrix where i can put the list of the nodes as rows and columns and for every element of the matrix we put a 0 where we don't have an edge and 1 when we have a connection.



With this Adjacency matrixes is very easy to calculate the degree of a network since I can just sum up all the ones that i find in a determined row and find the degree of a determined node. The same process can be done for directed graphs, but we need to calculate the in and out degree instead.

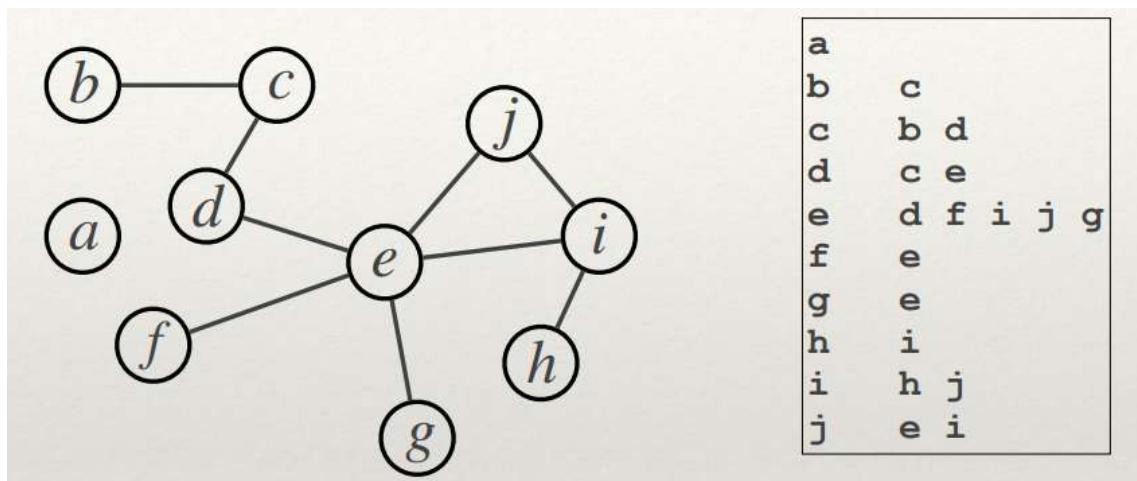
If we want to work with weights and calculate the strength we put the weights in the matrix and calculte the in-strength and the out-strength accordingly.

The bad part of this representation is that in real life we work with mostly sparse memory, and this matrix occupies NxN space inside of out memory, but we don't use all this space!

9 SPARSE NETWORK REPRESENTATION

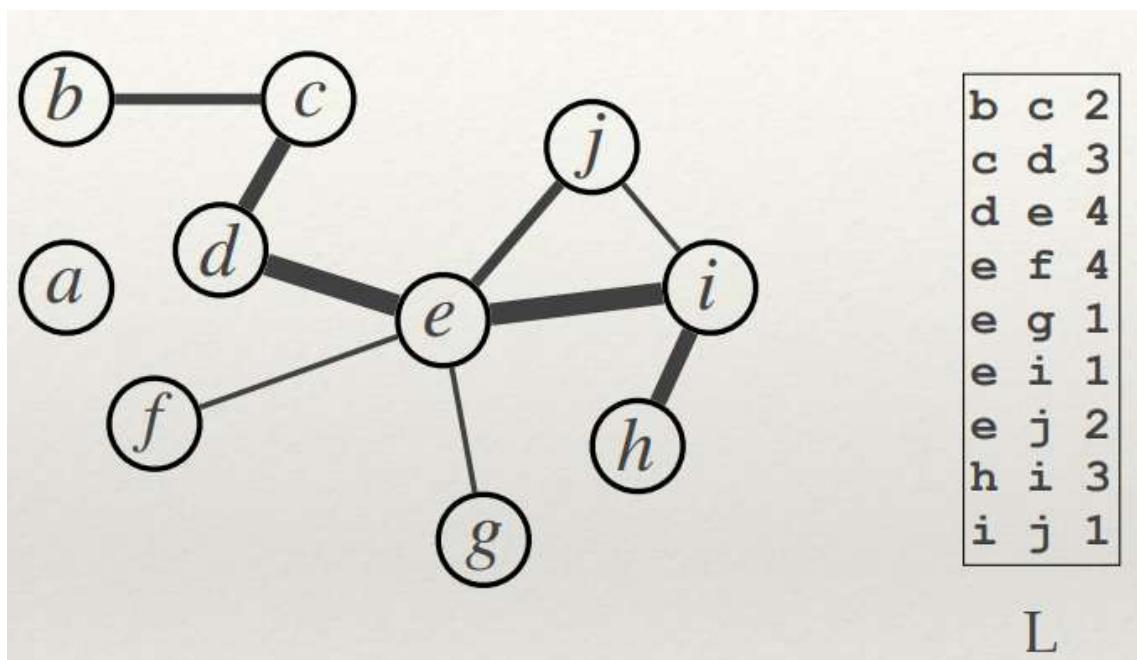
The storage needed by an adjacency matrix is N to the power of 2 and in sparse networks is highly inefficient since most of the space is wasted storing zeroes in the matrix and for large networks this is undoable. A solution to this problem, that is also way more efficient, is storing only the actual link and assume that if a link is not present in the list, then it doesn't exist. For this we can use adjacency lists or edge lists.

9.1 Adjacency Lists



Undirected network: list each link twice. Directed network: list only existing links

9.2 Edge Lists (with weights)



10 DRAWING NETWORKS

The problem of drawing networks is that we have a bi-dimensional space and we can plot all the nodes randomly in the space, but if we do this and we have many links between the nodes, our plotting would be very confused and un helpful. An idea is to display closely all the nodes that are close to each other and draw distant nodes far away in the plot. A network layout algorithm places nodes on a plane and visualizes the structure of the network.

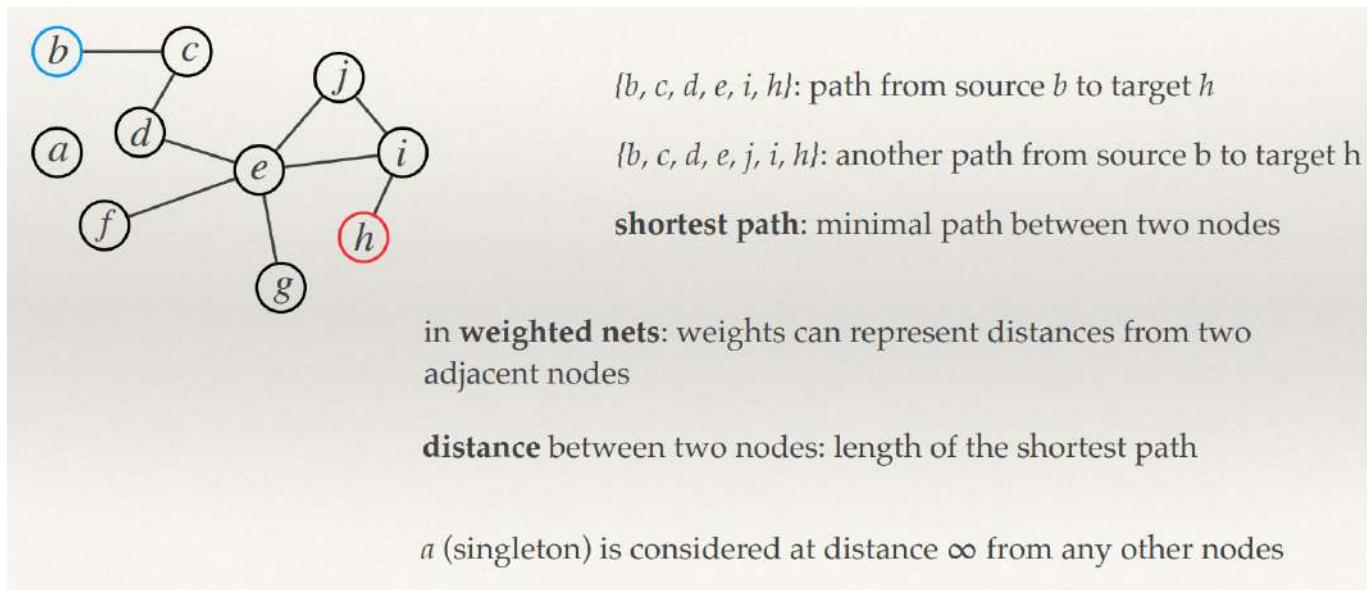
The ideas behind layout algorithms are to place connected nodes close to each other, to have links with similar length and to minimize link crossing and overlapping.

11 PATHS AND CONNECTIVITY

11.1 Paths

A path is a sequence of nodes ($n_1, n_2, n_3, \dots, n_l$) such that for every node i (n_i, n_{i+1}) is a link. The length of a path is l . If we have repeating nodes we have **cycles**, if we dont we have a **simple** path.

11.2 Shortest Paths



12 AVERAGE PATH LENGTH and DIAMETER

APL: Average Path Length

undirected graph: $\langle l \rangle = \frac{\sum_{ij} l_{ij}}{\binom{N}{2}} = \frac{2 \sum_{ij} l_{ij}}{N(N-1)}$

directed graph: $\langle l \rangle = \frac{\sum_{ij} l_{ij}}{N(N-1)}$

diameter of the graph: $l_{max} = \max_{i,j} l_{ij}$



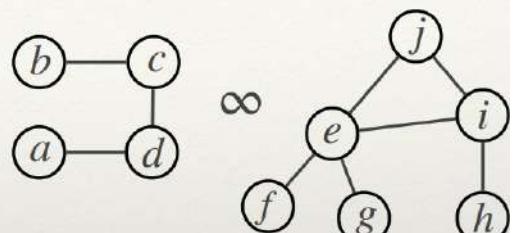
If we consider all the pairs of nodes in the network and we calculate their distance, we can compute the **average path length** $\langle l \rangle$. In an undirected graph we need to sum up all the distances between all the pairs ij and divide it by the binomial coefficient N chooses 2, while in the directed graphs we don't need to count all the edges twice.

The **diameter** of a graph (L_{max}) is the highest distance between two nodes, more precisely is the shortest path of the maximum length.

12.1 Average Path Length and Diameter with Disconnected Components

APL and diameter are undefined

$$\langle l \rangle = \frac{\sum_{ij} l_{ij}}{\binom{N}{2}} = \frac{2 \sum_{ij} l_{ij}}{N(N-1)} = \infty$$

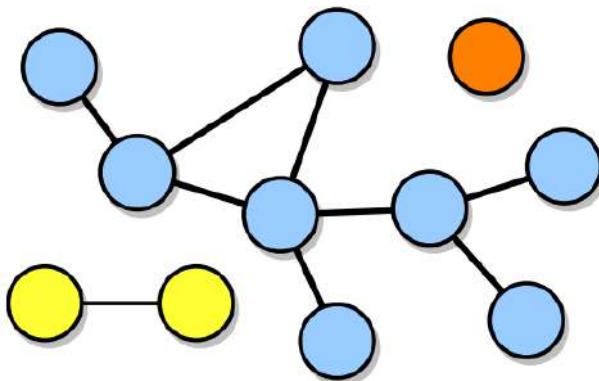


mathematical trick:

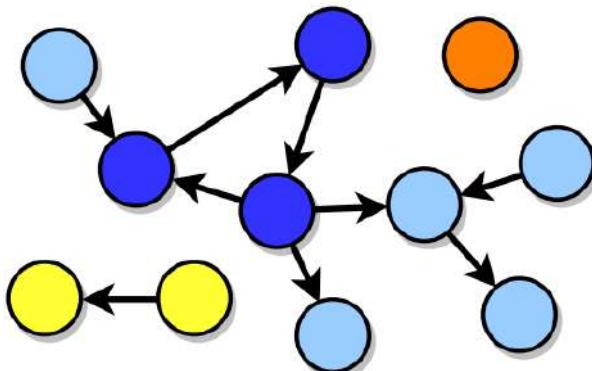
$$\langle l \rangle = \left(\frac{\sum_{ij} \frac{1}{l_{ij}}}{\binom{N}{2}} \right)^{-1}$$

In a graph where we have two connected components that are not connected with each other and hence are not reachable. In this situation we cannot define the APL or the diameter of the network, but we can calculate the sum of the inverse of the distance ($1/l_{ij}$) and divide it by the binomial coefficient and we compute the inverse of this equation. If we have only one component in our network we have the same result, but if we have two or more components we find that the distance between two not connected nodes is 0 (and not infinity) and we can still calculate the APL

13 CONNECTEDNESS AND COMPONENTS



A network is **connected** if there is a path between any two nodes, and if a network is not connected it is **disconnected** and has multiple connected components. A **connected component** is a connected subnetwork, and if we filter out all the other non connected nodes we find a connected network. The largest connected component is called **giant component** and is usually the most important one where studies are conducted (because the contagiousness of a virus is higher in big networks). A **singleton** is the smallest possible connected component.



In directed networks things vary a little. We have a graph that can have disconnected components, singletons, etc.. but if we look at connected components and we have a path from a source node to a target node it is not said that we have a path back from the target node to the source one. If we don't have cyclic paths between connected nodes we call them **weakly connected components**, while if we do we call them **strongly connected components**. The **in-component** of a strongly connected component S is the set of nodes from which one can reach S, but that cannot be reached from S directly. The **out-component** of a strongly connected component S is the set of nodes that can be reached from S, but from which one cannot reach S directly.

14 SMALL WORLD PHENOMENON

This is an experimental result that was observed when it was needed to know how many grades of separation where present between persons that don't know each other.

14.1 Milgram's Experiment

Milgram selected in a random way some persons of the midwest and requested that they send a letter to someone that they don't know that lives in Massachusetts. All of these people know the name of the recipient, his profession and where he lives (but not the precise address). Each of the participants was requested to sent their letter to a person that they knew that, according to their knowledge, had a higher chance to know the final recipient. The other person would then do the same until the letter reaches the final recipient. At the end of this experiment he observed that, in average, 6.5 steps were required to reach the final destination, hence we have the **Six Degrees of Separation**, that indicates that there are 6 steps between people that don't know each other. This was then called the **small world phenomenon** and indicates that the world is actually smaller than we may think.

14.2 When a Path is really short?

It depends on the size of the network. We have a relation between APL and network size when we consider networks of different sizes. We can determine that the APL is short when it grows very slowly with the size of the network, we can say then that the APL grows logarithmically.

15 A FRIEND OF A FRIEND - CLUSTERING COEFFICIENT

We can be in situations where we meet someone, and then this person knows someone that we know. In social networks we have triangles that indicate relationships, to say, if we take three nodes A,B,C if A is friend with B and C, then there is a high chance that sooner or later B and C become friends as well. To measure this chance we use the **Clustering Coefficient** of nodes, that is the fraction of pairs of the node's neighbors that are connected to each other.

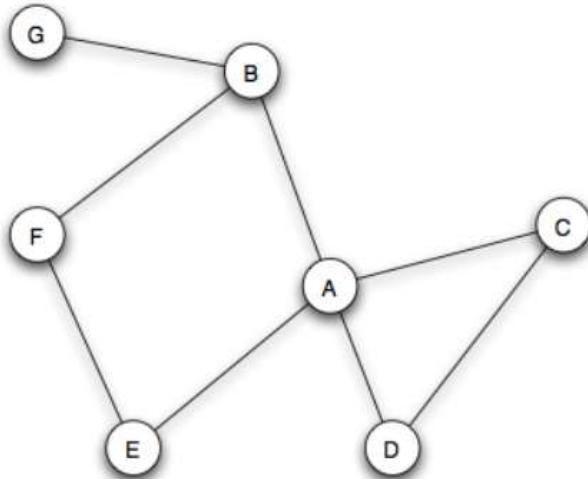
If N_i is the set of neighbors of i and $\tau(i)$ is the number of triangles involving i , we can count all the triangles that involve i and divide it by the maximum number that can involve i . K_i is the degree of a certain node, so the number of its neighbors.

$$C(i) = \frac{\tau(i)}{\tau_{max}(i)} = \frac{\tau(i)}{\binom{k_i}{2}} = \frac{2\tau(i)}{k_i(k_i - 1)} \quad : |N_i| = k_i$$

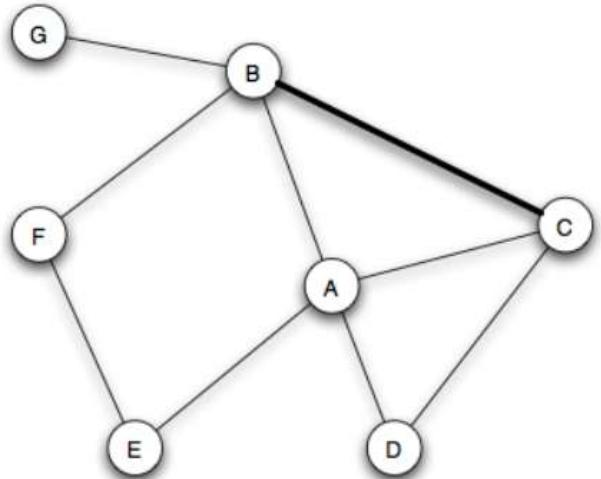
We can calculate the **Average Clustering Coefficient** of a network computing the average of the single coefficients of all the nodes. Social networks tend to have high clustering coefficient due to **triadic closure** (we meet friends through common friends)

16 TRIADIC CLOSURE (Intro to Strength of Weak Ties)

Friendship happen at a local level, I don't have an idea of the topology of the network and I don't know how my friends can change this topology but this relationships that are formed on a local basis have an impact on the global structure. The clustering coefficient and the characteristics that triangles are easy to be closed, can have effects on other nodes that can propagate information widely on a global level.



(a) Before B-C edge forms.



(b) After B-C edge forms.

We need to understand the mechanisms that govern how nodes come and go, and consequently how edges come and go. The answer to this is that if two people in a social environment have a friend in common, then there is an increased chance that they will become friends themselves at some point in the future, and this is the **Triadic Closure**.

The role of triadic closure in social environment has created the need to formulate measures to capture its prevalence, like the clustering coefficient. The clustering coefficient of a node A can also be defined as the probability that two randomly selected friends of A are friends with each other.

If we analyze a network for an extended period of time we can see the formation of edges, some through triadic closure and some not. The more strongly triadic closure is operating in the neighborhood of the node, the higher the clustering coefficient will tend to be.

A high clustering coefficient is frequent in social networks and hence we have an high friendship transitivity. The idea is that our network is made of N nodes and L edges, and we can take the same nodes and edges and re-shuffle them randomly so that we create a random analog of our real network. We then calculate the clustering coefficients in both networks and if the clustering coefficient in the real network is higher then the random one, we can then say that the clustering coefficient is higher then expected, since the cc is seen as the chance that my neighbors are linked with each other.

16.1 Reasons For Triadic Closure

- **Opportunity:** if A spends time with both B and C, then there is an increased chance that they will end up knowing each other and become friends
- **Trusting:** B and C are friends with A and they have a basis for trusting each other that a pair of unconnected people might not have.
- **Incentive:** if A is friends with B and C, then it becomes a source of annoyance (latent stress) in these relationships if B and C are not friends with each other. Even if i have a lot of friends but a low clustering coefficient, the latent stress might lead me to commit suicide as a study has shown. The structure of social connections can be interpreted as an indicator to prevent some catastrophic events.

17 THE STRENGTH OF WEAK TIES

Formulated by Granovetter in 1973 and he interviewed people who had changed jobs lately and wanted to know how they discovered this jobs. He found out that many people learned the information that led to their new jobs through personal contacts and these personal contacts where described as acquaintances rather then friends during the job interview.

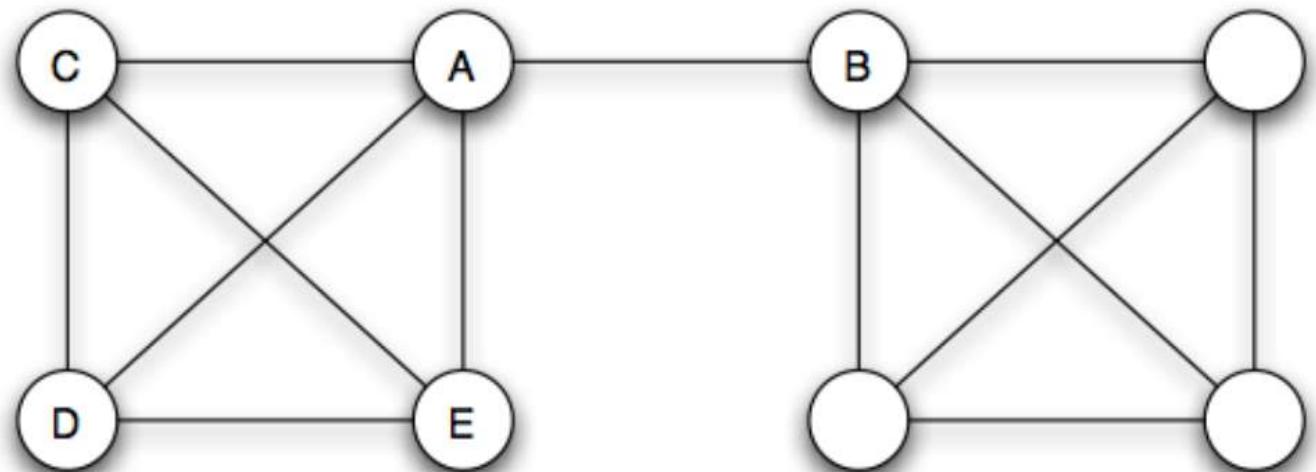
If we think about people, they are linked with each other through strong or weak ties (if i call someone everyday i have a strong tie, if once a year is just a weak tie). Apparently weak ties are more useful then strong ties. Granovetter formulated two reasons as to why weak ties used for finding relevant information are more efficient:

- A **structural** one that is based on how different friendships span different sections of the full network
- An **interpersonal** one that considers the purely local consequences that follow from a friendship between two people being either strong or weak.

Triadic Clouse is one of the crucial factors that are needed to define the strength of the links (if they are weak or strong). We can say that a weak tie is just somebody I meet sporadically and a strong tie is a friend i see often.

17.1 Bridges

Let's say that I find out from a friend of mine that a company proposes a dream job, I find myself interested in this topic because I see an **opportunity** and I seek an **incentive**. Let's say that i'm B and my friend is A and C,D,E are the employers, I can see that my friend has access to information that I don't have.

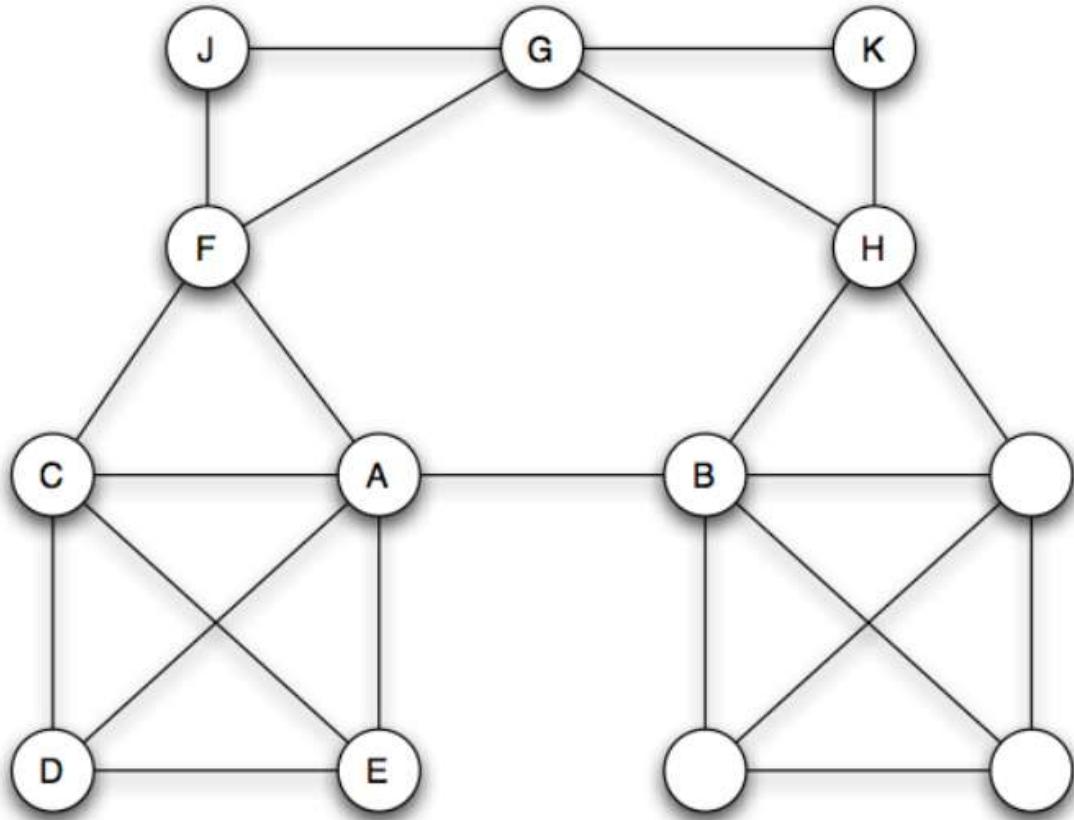


A, C, D, E are part of a cluster that is tightly connected with a high cc, while A and B are different since they have a lot of neighbors each but they don't share none of them, and this links are called **bridges**.

If I remove a bridge I obtain two different disconnected components. This bridge is the only route between two endpoints A and B. In social networks bridges like this are extremely unlikely since people tend to know each other. You may have a friend from a different background, and it may seem that your friendship is the only thing that bridges your worlds, but one expects in reality that there will be other, hard-to-discover and hidden paths that also connects these worlds.

17.2 Local Bridges

We need to introduce **Local Bridges**, if we take two nodes A and B in our network and we remove the AB direct link, we cannot isolate A,B from each other. This means that somehow A and B can still connect and communicate with one another through a longer path. Local bridges are much more common than a normal bridge in real social networks.



We say that an edge joining two nodes A and B in a graph is a local bridge if its endpoints A and B have no friends in common, more precisely deleting the edge would increase the distance between A and B to a value greater than 2.

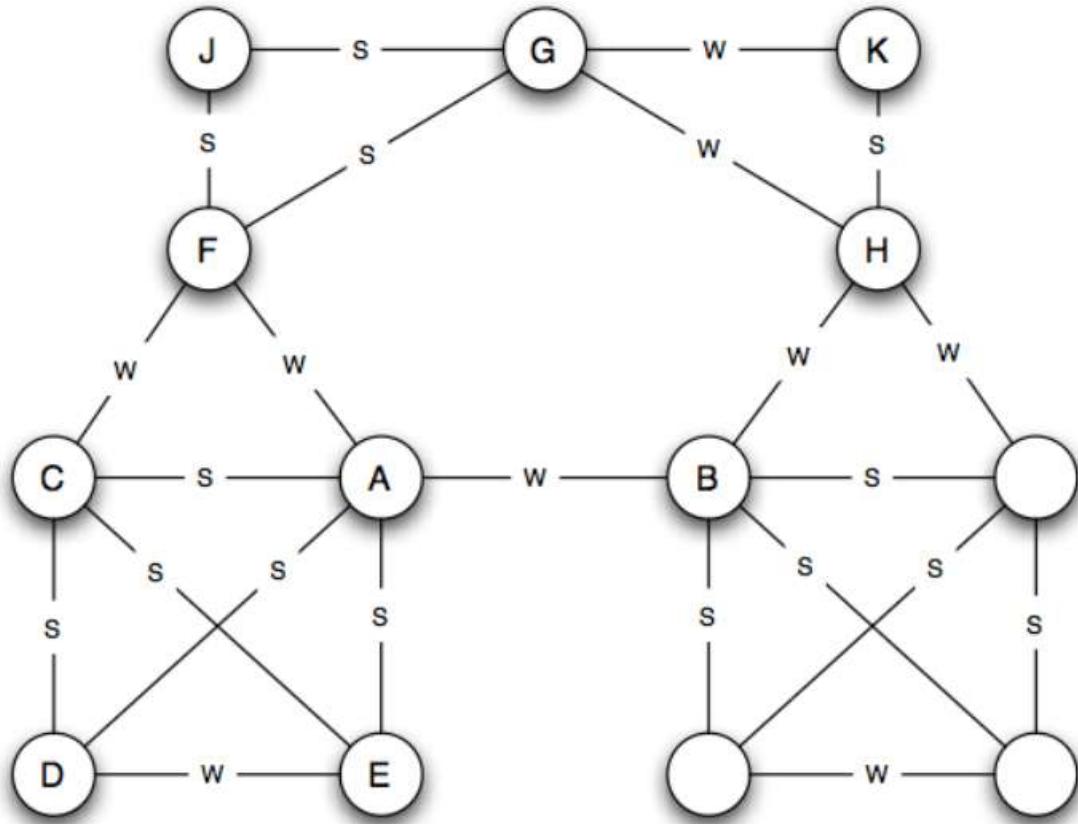
The **span** of a local bridge is the distance its endpoints would be from each other if the edge were deleted. The definition of a local bridge makes an implicit connection with triadic closure, since the two notions form conceptual opposites: an edge is a local bridge precisely when it does not form a side of any triangle in the graph. Where I have a bridge I cannot have a triadic closure and viceversa.

Local bridges, play a similar role to bridges, since they provide their endpoints with access to parts of the network, and hence sources of information, that they would otherwise not be able to gain.

The closely-knit groups that you belong to, though they are filled with people eager to help, are also filled with people who know roughly the same things that you do.

17.3 Different Level of Strength

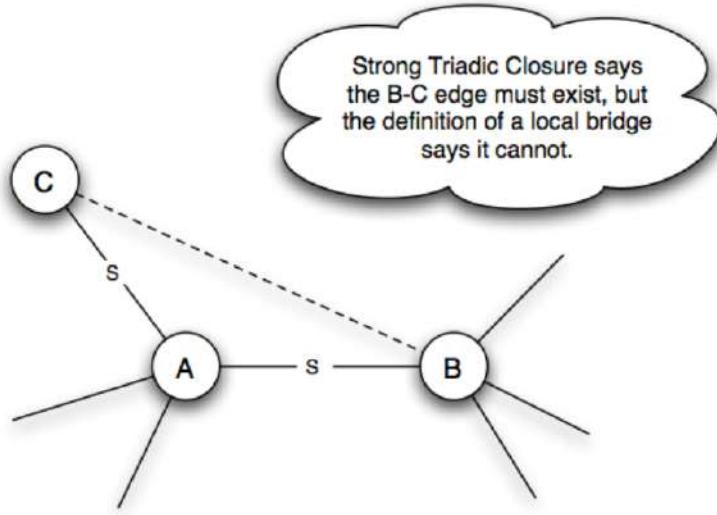
Stronger links represent closer friendships and higher frequency of interactions.



We want to prove that if A and B are connected by a local bridge then their strength must be weak, so we are linking the idea of strength of a tie (that has to do with interpersonal relationships) with a structural concept (idea of bridge).

17.4 Strong Triadic Closure

We will use in this case a qualitative assumption of the triadic closure: *If a node A has edges to nodes B and C, then the B-C edge is especially likely to form if A's edges to B and C are both strong ties.*



We say that a node A violates the Strong Triadic Closure Property if it has strong ties to two other nodes B and C, and there is no edge at all (either a strong or weak tie) between B and C. We say that a node A satisfies the Strong Triadic Closure Property if it does not violate it.

The idea is that if I'm a close friend of two persons that I meet frequently, it is very likely that my two friends are friends with each other. The Strong Triadic Closure property is too extreme to be expected to hold across all nodes of a large social network, so it is often used as an abstraction.

17.5 Local Bridges and Weak Ties - Proof

If a node A in a network satisfies the Strong Triadic Closure Property and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie. Simplified, if the two friends (B and C) don't have a link, then the link between A and B must be a weak one, and a bridge. If the link between A and B is a bridge then this link is a weak tie.

The proof is given by **contradiction**.

Consider a node A that satisfies the Strong Triadic Closure Property and is involved in at least two strong ties. Now we suppose that A is involved in a local bridge — say, to a node B — that is a strong tie. First, since A is involved in at least two strong ties, and the edge to B is only one of them, it must have a strong tie to some other node, which we'll call C.

But is there an edge connecting B and C?

Since the edge from A to B is a local bridge, A and B must have no friends in common, and so the B-C edge must not exist. This contradicts Strong Triadic Closure: since the A-B and A-C edges are both strong ties, the B-C edge must exist. **This contradiction shows that the existence of a local bridge that is a strong tie, cannot hold.**

This proof completes the connection between the local property of tie strength and the global property of serving as a local bridge. It gives us a way to think about the way in which interpersonal properties of social-network links are related to broader considerations about the network's structure.

These concepts work pretty well even in the real world but we have to consider them as somewhat an approximation. We found out that strong ties are friendship ties and weak ties are fundamental to connect remote parts of a network.

18 STRENGTH TIES AND LARGE SCALE DATA NETWORK

After a whole lot of years after Granovetter's first work, these previsions hadn't been yet tested on large social networks due to the difficulty in finding data that captured in a trustworthy way the characteristics of the links in a large network.

But as soon as digital communication was born, this data has somehow become available. The **Who-talks-to-whom** type networks were born in which we find all the ingredients we need: nodes, ties and strength of the ties (it is in fact assumed that the more two nodes A and B communicate, the stronger their tie is).

Some important case studies:

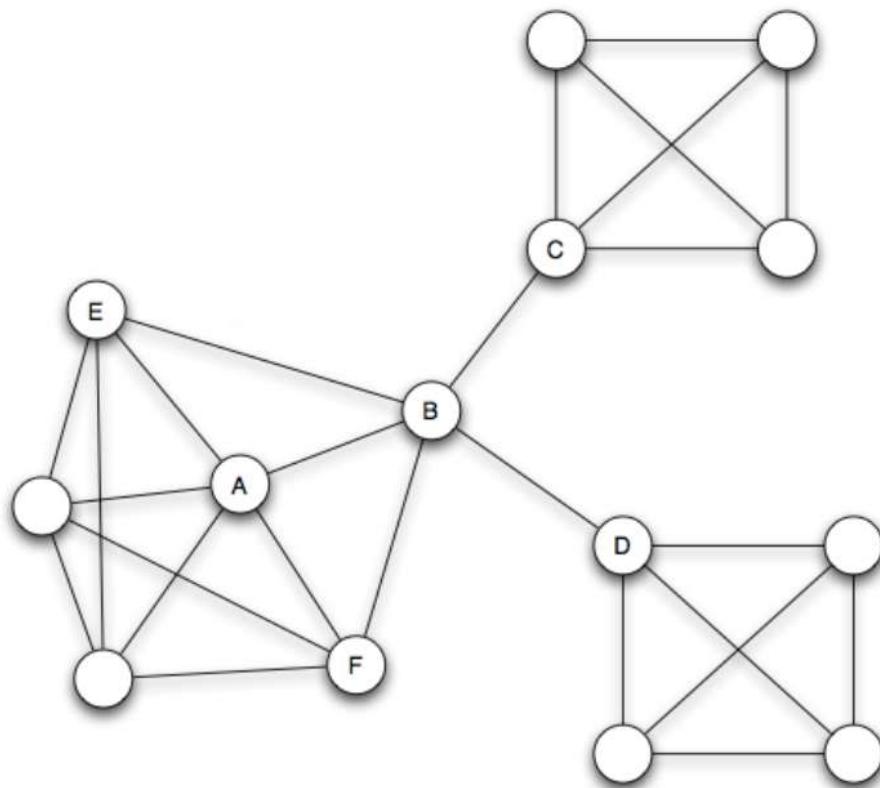
- **Cellular Network:** 2007, almost 20 percent of the whole population was covered. We have a massive Giant Connected Component where 82 percent of the nodes are in this component.
- **Indirect work of Onnela et. Al** where they got a network and they started removing the ties one by one. When they started removing ties starting from the strongest to the weakest, the giant connected component broke down pretty slowly, while if we removed starting from the weaker ties, the whole network collapsed faster.
- **Analysis of Facebook done by Cameron Marlow.** They created three categories of links:
 - **mutual:** if the user both sent messages to the friend at the other end of the link, and also received messages from them during the observation period.
 - **one-way communication:** if the user sent one or more messages to the friend at the other end of the link
 - **maintained relationship:** if the user followed information about the friend at the other end of the link, whether or not actual communication took place;
- **Analysis of Twitter** done by Huberman, Romero and Wu, where they confirmed that strong ties are very few for the majority of the users.

A fundamental concept for social networks is the **passive engagement** which means to feed social information to a person without a real communication. To put it more frankly: the likes, shares, visuals and etc.. is the new way of communicating.

Since we usually use indirect measures we can't really grasp the context and the true meaning of the real relationships between the nodes. Unless we have the complete control of the network we can't really get much from the analytics that we run.

Since the weak ties are so important, sometimes sections of the networks can result faster and even more used then other. Trivially strong ties are less convenient since to build a strong relationship I'm required to put a lot of time into it.

18.1 Contrast between Nodes



To keep things clear, we can say that certain nodes are more important than others, hence we find the concept of **Heterogeneity**, and we can have certain nodes that function as links between other components. The more interconnected node (A), that has a higher cluster coefficient and an higher involvement in triadic closure and this characteristic is called **Embeddedness**.

18.2 Misbehavior

The presence of mutual friends puts the interactions between two people “on display” in a social sense, even when they are carried out in private. In the event of misbehavior by one of the two parties to the interaction, there is the potential for social sanctions and reputational consequences from their mutual friends. No similar kind of deterring threat exists for edges with zero embeddedness, since there is no one who knows both people involved in the interaction.

18.3 Structural Hole - Social Capital

If we remove certain edges (bridges like BC or BD) we get as a result some disconnected components. This situations is called **Structural Hole**. A node in this position offers advantages in several dimensions to relative nodes with high embeddedness, and one of this advantages is the **Informational advantage**: such a node has early access to information originating in multiple, non-interacting parts of the network. This node is investing her energy efficiently by reaching out to different groups rather than basing all her contacts in the same group.

We can also quote the concept of **Social Capital**, which means to take advantage of our own social position inside a network to gain advantages, an example nowadays could be the work of influencers.

19 HOMOPHILY

Homophility is translated in a popular saying that is that *birds of a feather flock together*, which in italian is translated in *dimmi con chi vai e ti diro chi sei*. The idea of homophily is that is very likely for me to be connected to people that are similar to me.

The surrounding context of homophily requires us to focus on:

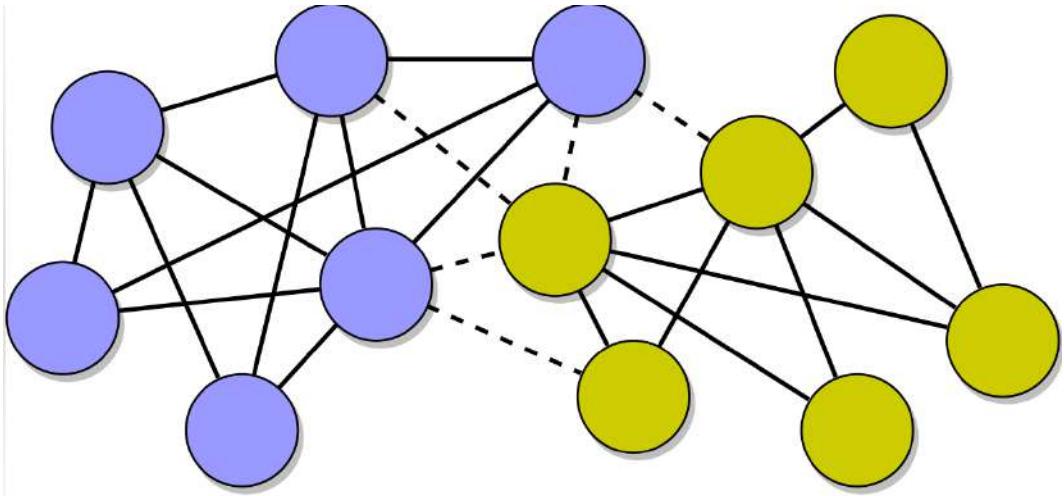
- **structural properties of networks**: how nodes are linked to each other
- **link formation processes**: how links are formed, and this allows us to introduce predictive models that allow us to predict which links are going to be formed before others
- **personal characteristics**: we link structural properties to personal characteristics (like the bridge and the weak ties) and with this we can get at the end **similarities between individuals**.
- **surrounding contexts**: this are the factors that exist outside the nodes and the edges of a network but can have an effect on the evolution of the network itself.

19.1 Definition of Homophily

Is the principle according to which we tend to be similar to our friends. If we meet someone randomly outside of our social circle we can see that with an high chance that person is quite different from me. This translates in saying that your friends are not statistically significant as a random sample of the population, if we take into consideration an idea that me and my friends have and that is different with the opinion of the outside world, our opinion may be flawed. This happens because since me and my friends share similarities in multiple fields, and I cannot pretend that my friends make a statistical significant sample of the population, even if I have a lot of friends.

If we want to reason according to similarities on individuals we need to come to an agreement. Each individual has some **immutable characteristics** (like the ethic group I belong to, etc..) but there are also **mutual characteristics** (like interest in things such as music, films etc..).

19.2 Birds of a Feather Flock Together



We can have a certain trait (color) that characterizes some nodes and we want to count the number of links **intra-groups** between the nodes of the same color, and the number of **inter-group** link that connect nodes with different traits. Structurally speaking, **the idea is that if I observe the number of cross-edges (inter-group) is much lower than the intra-group edges then we have homophily.** This means that it is much more likely for two similar node to be connected to each other compared to two different nodes.

It's just a matter of probability, we need to calculate the frequencies of links (intra-groups and cross-groups) and we need to compare this frequencies with the probability to observe such number of links inside of a random network.

19.3 Measuring Homophily

To measure homophily we have to run 3 steps:

- assign a random color to each node
- count the number of cross-color edges
- compare the numbers with the actual network.

The idea is that we have a network that has been created from real data, if we want to understand if a feature is characterizing that network we compare a measure calculated in that network with the same measure calculated in a randomly generated (but analogous) network.

Measuring Homophily

- Total edges = 18
- Right nodes = 3, Left nodes = 6
- Cross-group edges = 5
- Compare this network to a randomly formed network with 18 edges

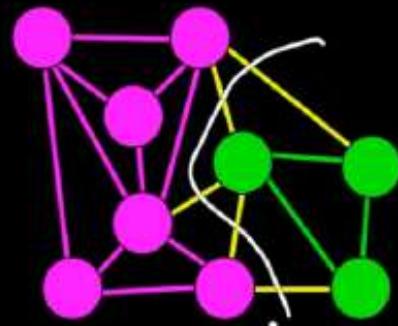


Figura 25: Esempio di omofilia, con il calcolo del rapporto

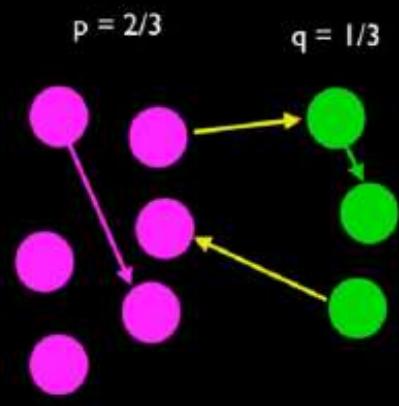
Random Link Construction

$$\text{Probability of a p-p link} = 2/3 * 2/3 = 4/9$$

$$\text{Probability of a p-q link} = 2/3 * 1/3 = 2/9$$

$$\text{Probability of a q-q link} = 1/3 * 1/3 = 1/9$$

$$\text{Probability of a q-p link} = 1/3 * 2/3 = 2/9$$



$$(p + q)^2 = p^2 + 2pq + q^2 = 1$$

Figura 26: Esempio di omofilia, con metodo del random link construction

Once we assign the colors (p and q) inside our network, we estimate the number of links between p -color nodes, and that is given by $p*p$ which is $4/9$. The number between q -color nodes is the same $q*q$ and is $1/9$. The number of cross edges is just $p*q$ but we have to count this twice since it is an undirected network, so it is $2*p*q$, and this is an estimation of the number of links between nodes of different colors ($4/9$). Now we calculate the number of links between p -color, between q -color and between p -color and q -color, and we see that we have a total of 18 edges, and 5 cross-edges between p -color nodes and q -color nodes. If we compare the estimation of cross-edges given by $2*p*q$ ($4/9$) with the actual number of edges ($5/18$) we see that $4/9 > 5/18$ and this means that the actual number of cross-edges is less than the number of expected edges across different groups.

Measuring Homophily

When the fraction of actual cross-group links are significantly less than $2pq$, then there is evidence of homophily

$$q = 6/9, p = 3/9 \quad 2pq = 4 / 9 = 0.44$$

$$\text{Actual edges} = 5 / 18 = 0.27$$

0.27 is "significantly less than" 0.44

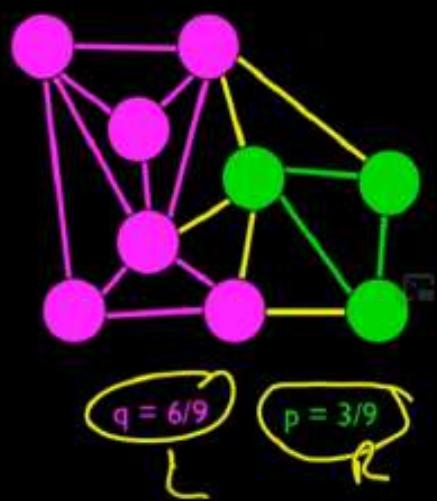


Figura 27: Misura di omofilia finale

The homophily test is done by just checking if the number of actual cross group edges is less than $2pq$, where p is the number of p -color nodes in the network, and q of q -color nodes. If the actual number is less, then we have homophily. More precisely, if the fraction of cross-types edges is significantly less than $2pq$, then there is a **signal of homophily**.

We can have perfect homophily when we have no edges across different groups (colors).

20 DEGREE ASSORTATIVITY

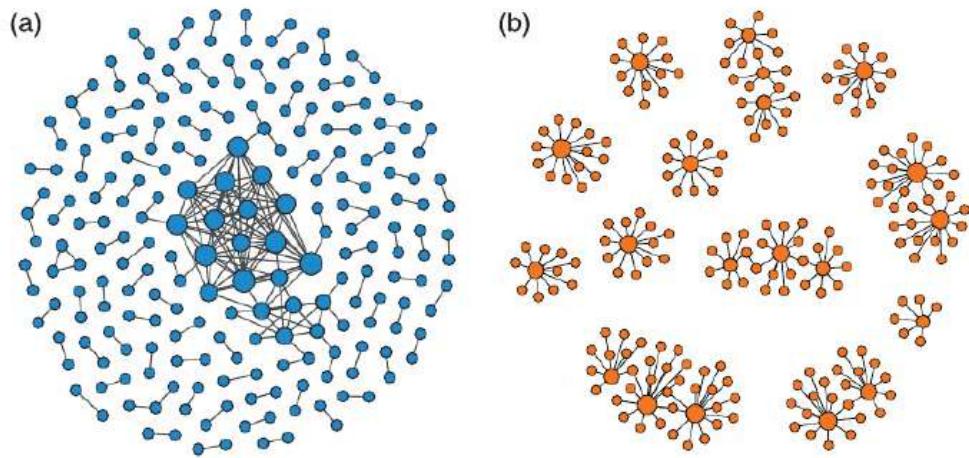


Figura 28: Esempio di reti assortative (a) e non assortative (b)

This concept is related with the structural properties of the network. Degree assortativity is just the homophily by degree. If we just consider one node, this node can have different characteristics and traits, but there are also structural characteristics for that node which are the number of connections of that node (the degree). So if we are looking for homophily, we can say that we are looking for also correlation between the presence of a link between two nodes and their similarities.

Hubs, which are nodes with higher degree with respect to the other, are more likely to be connected with each other, and this is called a **positive degree correlation**. The networks that have a core-periphery structure with hubs in the core are called **assortative networks**, like social networks.

Networks that have hub-and-spoke structures that are disconnected from each other are called **dissassortative networks**, like the web. In this situation we have a **negative degree correlation**.

This basic measure helps us to characterize in an high level order, our network.

20.1 Measuring Assortativity

A way to compute the degree assortativity is by measuring the **degree correlation function**, which is the correlation between the degree and the average degree of the neighbors of nodes with that degree. So if I have node i , the neighbor degree of i is $K_{nn}(i)$, this is easy to compute

Another way to compute the degree assortativity is by measuring the **degree correlation function**, i.e., the correlation between the degree and the average degree of the neighbors of nodes with that degree.

Let's calculate the average of i 's neighbors' degrees, first:

$$k_{nn}(i) = \frac{1}{k_i} \sum_j a_{ij} k_j$$

With a little abuse of notation, let's define the degree correlation function as the average of all the $k_{nn}(k_i)$ for all the nodes whose degree is k

$$k_{nn}(k) = \langle k_{nn}(i) \rangle_{i: k_i=k}$$

to check visually network assortativity, we plot:

$$(k, k_{nn}(k))$$

if we consider the adjacency matrix that describes the graph. We compute the sum of $A_{ij}K_j$ for all the j 's that are neighbors of i . This is then divided for the number of neighbors ($1/k_i$). A_{ij} can be 0 or 1 according to the presence or not of links between i and j .

We define then degree correlation function as the average of all the $K_{nn}(k_i)$ for all the nodes whose degree is k and not only for node i , and then we calculate the average for all the $K_{nn}(i)$ for which the degree of i (k_i) is equal to k . We consider all the nodes whose degree is k and for all this nodes we calculate the average of their neighbors degree and the total average is $K_{nn}(k)$.

To check this visually we just plot k and $K_{nn}(k)$.

20.2 Types of Assortativity

We have three types of Assortativity

- **positive:** we have a centre and a periphery, like in a scientific collaboration network because we collaborate with the whole community but hold relations with a small part of these groups
- **neutral:** we don't have a precise structure like in power grids
- **negative:** we have a lot of small components sparse like hub-and-spokes and stars

21 SELECTION AND SOCIAL INFLUENCE (Underlying Homophily)

We have two main aspects from which homophily and assortativity can emerge:

- **selection**: when similar nodes get connected with each other, like the birds that flock together
- **social influence**: connected nodes become more and more connected with each other

Selection and social influence are different directions of the cause-effect relationship that makes homophily a thing that can be measured and observed in networks, because in one occasion we have that two nodes are connected because they are similar, and in the other they start agreeing more and more about something.

Put like this can be a problem since our social circle become composed of people that are much alike, and this is considered to be one of the triggers of the **echo chamber**, where our circle si not diverse and out opinions get stronger and stronger by the passing of time and we are not exposed to other opinions that could help us. We consider people outside our circle like outsiders and we don't trust them, hence not considering different opinions from ours.

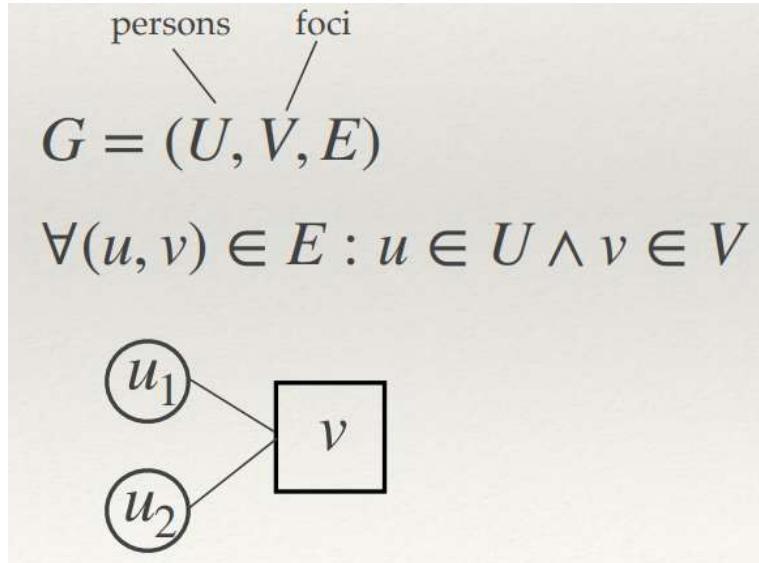
21.1 Interplay between Selection and Social Influence

Case studies are very hard to be analyzed. Two persons that share the same interests become friends and after a while they start sharing other interests and after a long period of time from a network-visualization point of view we see only that these two persons are connected, **but we don't know for what reason they connected**.

If we want to observe both the factors we need to take samples of the datasets for extended periods of time (years), and here we understand that the main difficulty of these studies is that cause-effect relationships is difficult to grasp since we don't know if selection comes before social influence or viceversa.

If we want to study complex phenomena like drug addiction we need to perform this type of analysis. Am I an addict because of my social circle or since I'm an addict I find my self better with other addicts? Or is it **peer pressure**: I belong to a social circle where everyone starts taking drugs and I start myself.

22 AFFILIATION



We want to represent the surrounding contexts by the means of the network. We know how to represent relationships between two individuals connected to each other, but we want also to represent and connect individuals to an activity, an opinion or more precisely an **affiliation**. We have a general term to indicate this activities that are the **focal points** (FOCI), a representation for this relationships are bipartite graphs where we have a set of persons (U), a set V of focal points and the edges E . Every person u is connected to a focal point v only if they belong to different groups.

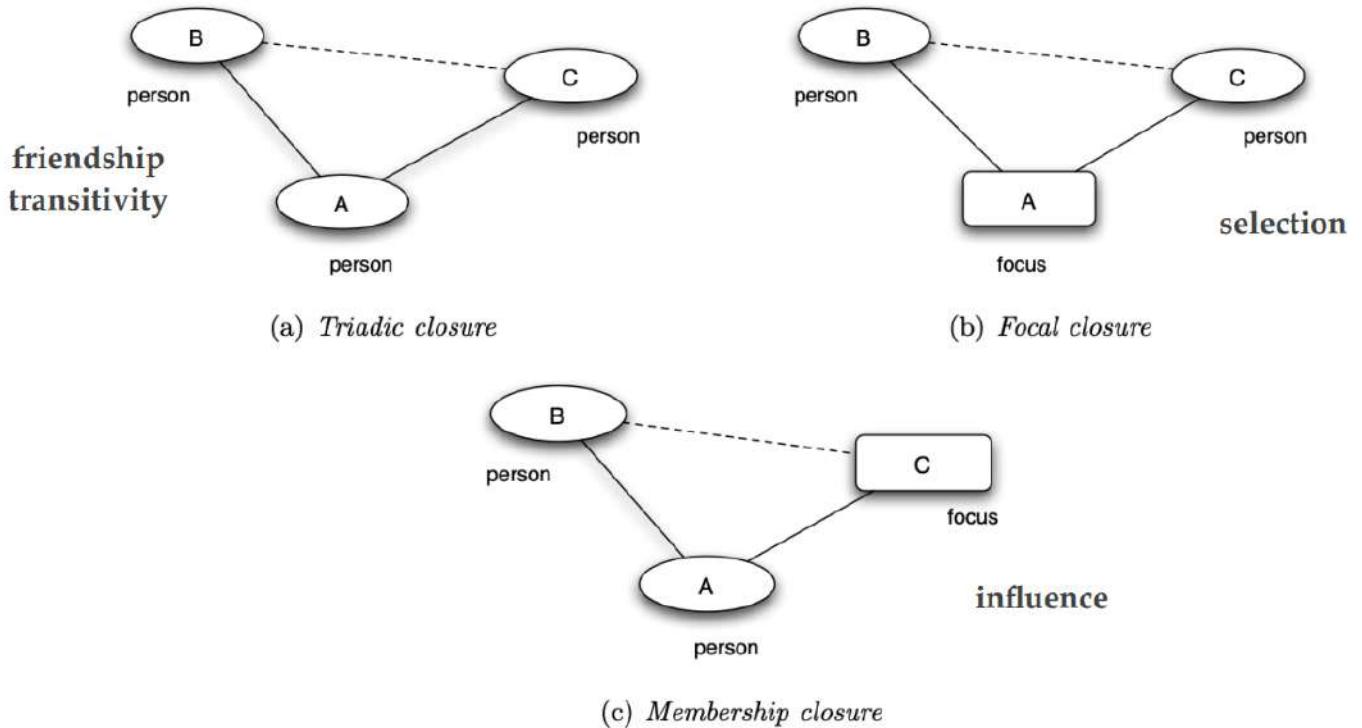
22.1 Affiliation Networks

We can complement our affiliation if in a situation two persons are affiliated to something like a karate club, and we can augment this affiliation network annotating friendship relationships between nodes. We can also use **projections** to connect people to predict something that can happen in social networks. If two people are connected to something like the board of a company, it is likely that these two people are connected to each other.

22.2 Co-Evolution of Social and Affiliation Networks

Both social and affiliation networks change over time and we want to merge the information from both networks into a **social-affiliation network**.

23 CLOSURES



- If A and B are friends and A and C are friends, it is possible through **triadic closure** that B and C become friends, and this is called **Friendship Transitivity**.
- If we look at a **Focal closure** we see that A is a focus (activity) and B and C are individuals that are connected to this activity, and we can predict that sooner or later B and C will connect within each other, and this is a closure that is gotten by **selection**, since they both attend the same activity.
- **Membership closure** is a little different, since we have two individuals connected by a relationship, and one of these two individuals is connected to a focus, so by **influence** we can predict that the other individual will be introduced in the future to the same focus.

We need to explain in a more precise and technical way how to compute this information. The idea to compute the probability of the formation of a link is the following:

- We proceed by taking temporal snapshots of the network at given times t and t'
- For each k we identify those nodes that have exactly k friends in common at the given time t but that are not yet connected to each other with a link, let's call this set X .
- If $T(k)$ is the fraction of the nodes that have connected and have formed an edge at the time t' .

It's pretty obvious that we need to consider this aspects on a probabilistic point of view, and so we need to define an empiric probability (a probability based on data gathered from a historical series, for which I can gather the relative probabilities) on the fact that an edge will be formed between two persons that have k friends in common, and how this probability varies in function of k .

Let's define p as the independent probability of forming a new edge between persons that don't know each other but that have a friend in common, and we assume that this probability is small. $1 - p$ represents the probability that this doesn't happen and is consequentially a larger number. If we put my k friends into the equation, the probability that with k common friends, the two persons will keep not forming an edge is $(1 - p)$ to the power of k .

From this formulation we get two **baselines**:

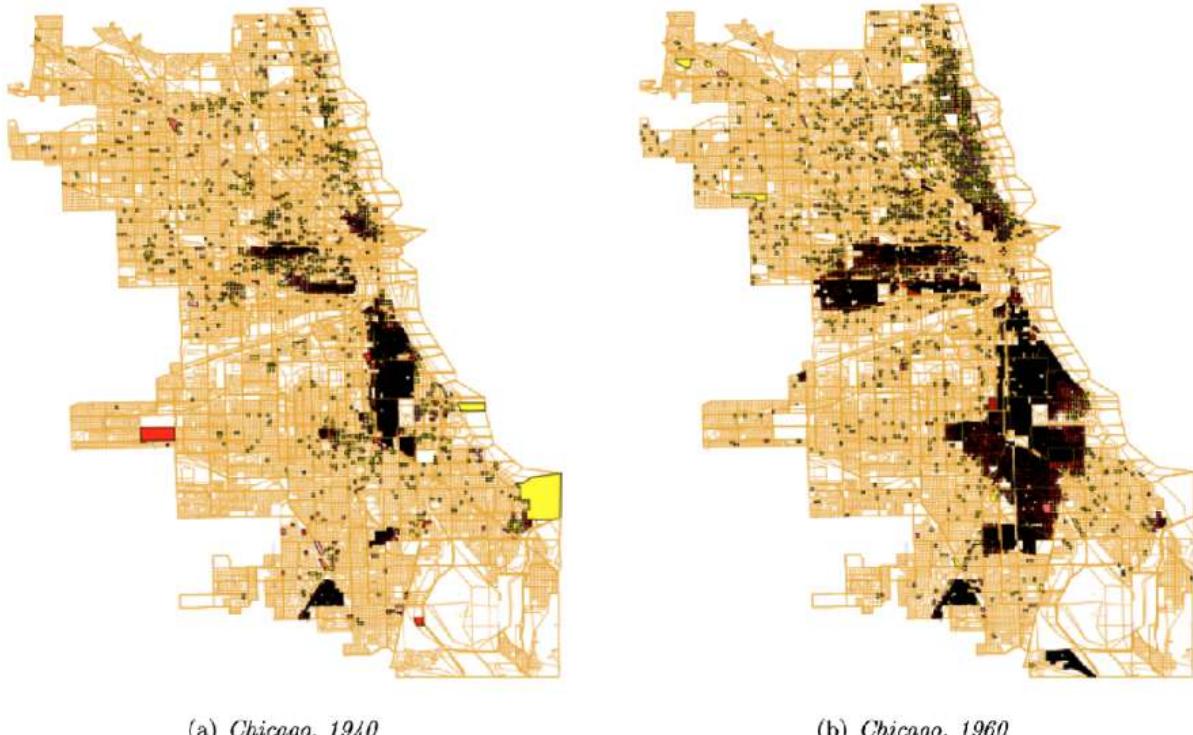
- $T_b(k)$ that is the complementary probability of $(1 - p)^k$, which is $T_b(k) = 1 - (1 - p)^k$
- We can compute $T_b(k)$ as the probability of a previous degree of k , kinda like with the Markov chains. We will then have $T_b(k) = 1 - (1 - p)^{k-1}$

We have some case studies worth mentioning:

- **The Kossinnetz and Watts Experiment:** they took the datasets of students emails and places they used to frequent and found out that the likelihood to form new friendships is proportional to the number of places in common between the persons.
- **Membership Closure on Wikipedia editors:** The probability to edit the same articles of my friends grows proportionally to the number of friends that already edited it.

24 SHELLING MODEL

24.1 Spatial Model of Segregation



(a) *Chicago, 1940*

(b) *Chicago, 1960*

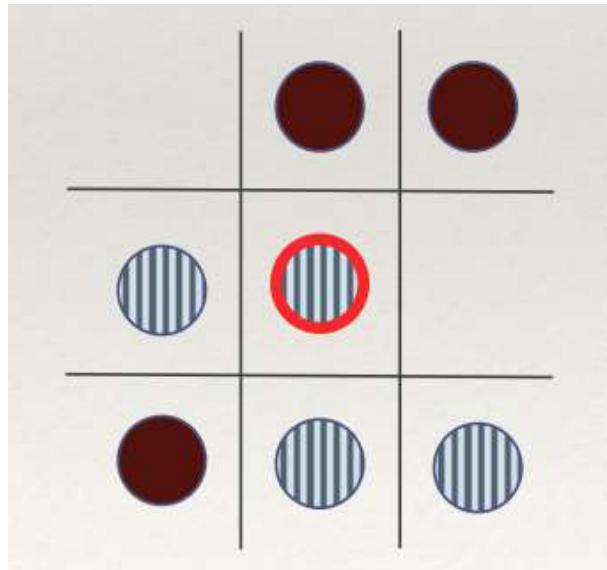
If we think about homophily, we can see the formation of groups of people that are tightly connected to each other and loosely connected to other distant groups. We can define this as segregation, since we have a subnetwork of people that is tightly connected inside the network itself but have very few connections to other groups, this could be caused by **friendship transitivity, selection or influence**. All these effects make these single groups more connected within the group, but also the individuals inside the groups start being much more alike to each other compared to other groups, by the **echo chamber** effect.

This happens in cities as well, if we take Chicago in a span of 20 years we can see that a certain ethnic group was dominant in a certain area, and the effect expanded in the next 20 years having the same ethnic affirmed in the same zone.

24.2 Formalization of Schelling Model

The question that motivated Schelling was: can spatial segregation arise from the effect of homophily operating at a local level?

The assumption is that no single individual wants to segregate explicitly, and Schelling wanted to explain this phenomena without considering racism as a cause for this phenomena. If we take into consideration two different types of agents that each have a different immutable characteristic. If we put these agents in the cells of a grid, so that each agent can have 8 different agents. Some of these cells are already occupied by other agents and some not. We can have up to 8 cells touching another agent, and we want to count how many neighbors are similar to me and how many are different from me.



Each agent wants to have at least t neighbors similar to him, if an agent finds less than t agents similar to him they want to move to another cell. I can stand up to a threshold of t people that are different from me, but after that I want to move.

24.3 The Dynamics of Movement

Agents move in sequences of rounds, if an agent is not satisfied, it moves in another cell of the grid. If there are no free cells around the unsatisfied agent we either move the agent randomly or just leave him where he is. The idea is to reiterate these agent dynamics over and over until we reach an equilibrium, our goal is to reduce the number of agents that are unsatisfied.

24.4 Interpretations of the Model

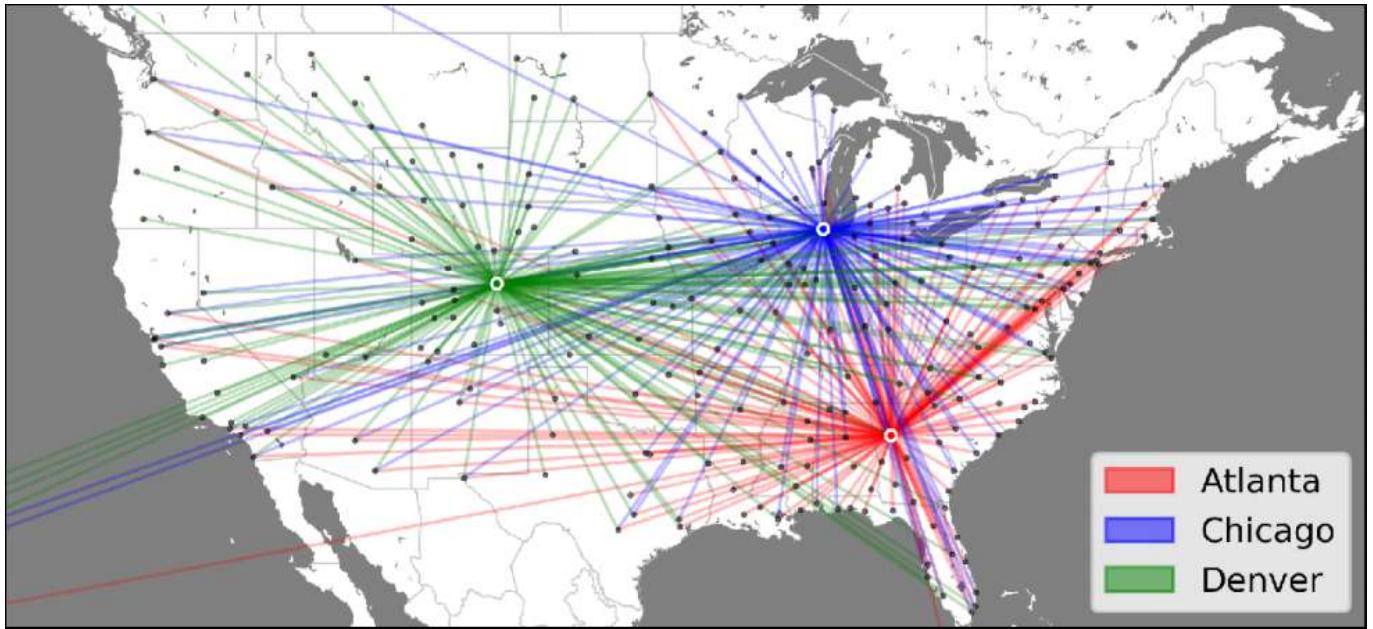
Segregation emerge as a natural phenomenon even when agents accept to be a minority. We can have a non-segregated configuration that satisfies our model, but the main point that Schelling makes is that these configurations are very unlikely. They can happen, but I need to design a certain city in a very specific way but if I leave the people to decide their own spot it's likely that even in a non racist environment, segregation would still happen.

If the t threshold is large, and hence we introduce racism in our society, the patterns get much more amplified and we don't only have segregation but even polarization where certain people live in zones where only people similar to them live. But there are cases where segregation comes out even with small thresholds.

25 NETWORK MEASURES

25.1 Real Networks are Heterogeneous

Many times we said that heterogeneity is a feature that characterized real networks in respect to toy problems that we study to understand mathematical bases of the measures. We need to asset that heterogeneity changes all the measures and operations that we can do with our data. We made the example of airports, and if we compare the degrees of three main airports to smaller airports we can see a great inequality in the distribution of links between the nodes, which are the airports. This makes us think that some airports (nodes) are more important than other, and we call them more central.



26 CENTRALITY MEASURES

The centrality is the measure of importance of a node. We have three different types of centrality measures: **Degree**, **Closeness** and **Betweenness**.

26.1 Degree Centrality

The degree of a node (K_i) is just the number of neighbors of that node i. Nodes that have a high degree are called hubs. A hub is a node whose degree is much higher compared to other nodes, and in order to have a scale for telling hubs apart from other nodes we need to define the average of a node of the network. To calculate the average we need to sum up the degree of all nodes and normalize this sum by N which is the total number of the nodes. $2L$ is the total number of links in an undirected network.

Average degree of the network:

$$\langle k \rangle = \frac{\sum_i k_i}{N} = \frac{2L}{N}$$

26.2 Closeness Centrality

If we think about centrality, it's very likely to think about some nodes whose distance to other nodes is similar to other, almost like the center of a circle, where we know that the distance from the center to the circle is the radius. **If we apply this way of thinking to networks we can get as a result that some nodes are more central than others, and closeness measures try to capture this type of information.** A node is more central the closer it is to the other nodes.

$$g_i = \frac{1}{\sum_{j \neq i} \ell_{ij}}$$

where ℓ_{ij} is the distance between nodes i and j

If we define the closeness of a node i as G_i we need to put at the denominator the sum of all the distances from node i to another node j , which is L_{ij} (apart from the node i itself) and we put 1 as the numerator, so its just the inverse of that sum.

If L_{ij} is the distance between nodes i and j , closeness is very low when the nodes are a large denominator, this means that the greater is the average distance with other node, the lower the value of closeness centrality is. And of course the higher this value is, the closer the node is to all the other nodes.

26.3 Betweenness Centrality

The idea is that a node is more central the more often it is crossed by paths. We can visualize this with the abstraction of a city, if we think about a central topological place like a square like piazza castello in turin, from this central point we can get to all the other points in the city is comparable and hence the average distance is low. If we think instead of streets or squares that are more important than others in the way of traffic control and congestion, we are not thinking about central point like piazza castello, but we start thinking about bridges and connections.

The idea of betweenness is that we have some nodes that are very important because they are often crossed by different flows of traffic. The mathematical definition of betweenness is just the sum of all the number of shortest paths ($\sigma_{hj}(i)$) from h to j that flow through node i divided by the number of shortest paths (σ_{hj}) from h to j.

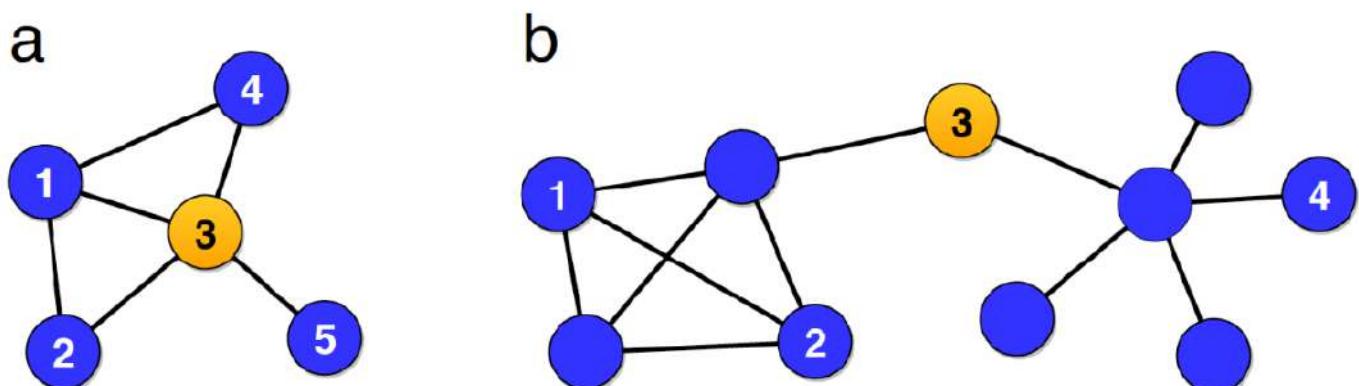
$$b_i = \sum_{h \neq j \neq i} \frac{\sigma_{hj}(i)}{\sigma_{hj}}$$

σ_{hj} = number of shortest paths from h to j

$\sigma_{hj}(i)$ = number of shortest paths from h to j running through i

If this number is close to 1 this means that the majority (if not all) of the shortest paths from every pair of nodes in the network flow through that specific node i. If we delete this node from the network, we get a split in the network. A node with an high betweenness in the network is critical not only for centrality in the topological interpretation, but for improving or deteriorating the quality of the traffic flow inside of the network.

Hubs usually have high betweenness but of course there can be nodes with high betweenness that are not hubs



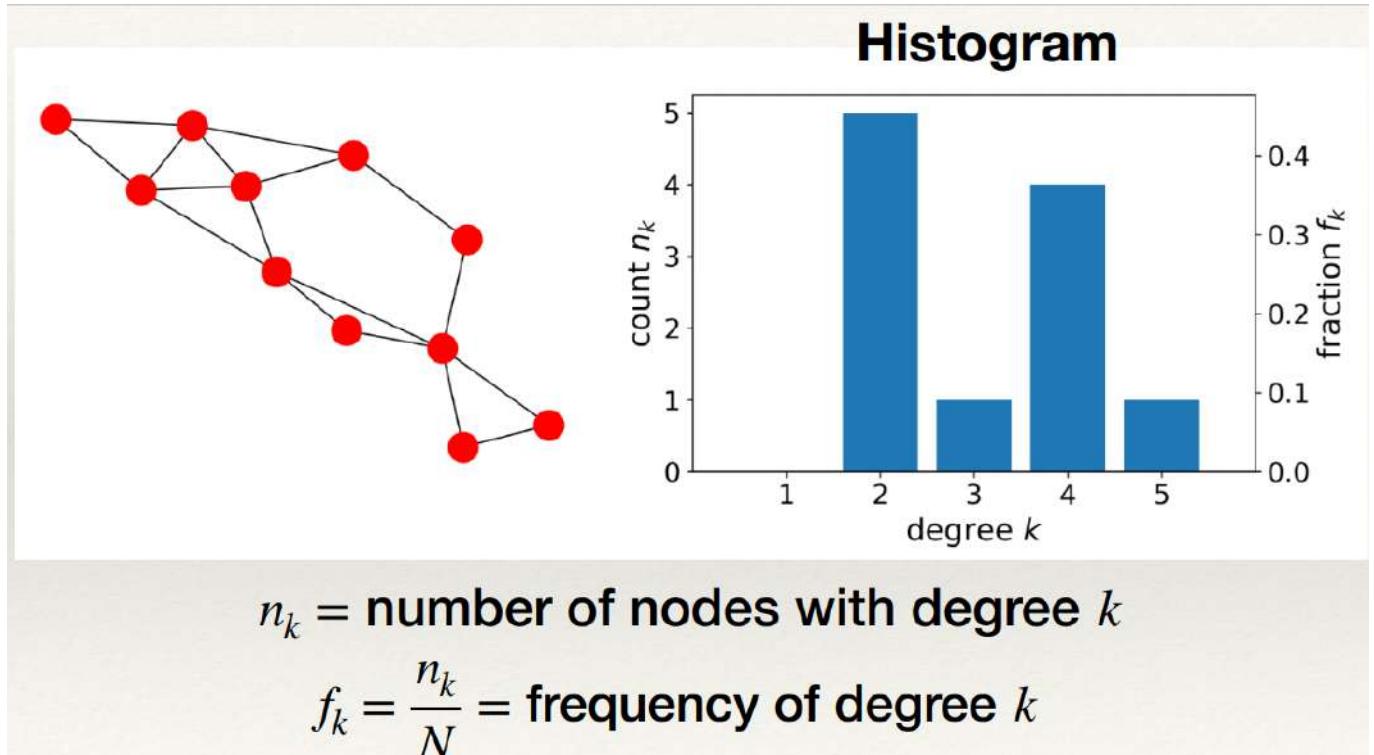
The concept of betweenness is quite similar to the concept of bridge. If we have a node with a high degree it doesn't necessarily mean that node has an high betweenness, but it can be important since it connects two parts of a network (b).

If we extend the definition of node betweenness to link betweenness we can calculate the **Link Betweenness**, which is the fraction of shortest paths among all possible node pairs that pass through the link.

27 CENTRALITY DISTRIBUTIONS

If we have a small graph like students in a class (toy example) it could be useful who is the hub, highest degree, closeness and betweenness in that example. We can then give a psycho-sociological role to every node in this network, but this is not enough in advanced network analysis.

In small networks it makes sense to ask who are the most important nodes, but in large networks it is not useful, and for this reason we need to have a statistical approach to this problem. The **statistical approach** that we can adapt in modern social network analysis is to focus on classes of nodes and links with similar properties, and we need to use distributions of such centrality measures.



If we have the degree on the x axis and we want to plot the number of nodes with that degree, I'm focusing on the class of nodes of degree 1,2,3,4... and this is just a representation of this distribution. If we want to compute the frequency of the nodes with degree k (F_k) we just divide the number of nodes with degree k (N_k) with the total number of nodes N .

We can value, empirically speaking, the frequency as a probability. If we assume that for large N , the frequency (F_k) becomes the probability P_k of having a degree of k , we can then plot the probability distribution of P_k versus k .

27.1 Cumulative Distributions

If the variable is not an integer like in the example of the betweenness, the range of the variable is divided into intervals called **bins** and from this we count how many values fall into these intervals, and this discretizations in intervals allows us to plot the distribution.

$$P(x) = \sum_{v \geq x} f_v$$

A cumulative distribution $P(x)$ is the probability that the variable takes values larger than x as a function of x . $P(x)$ is given by summing up the frequencies of the variables inside of the interval that are greater or equal than x .

27.2 Logarithmic Scale

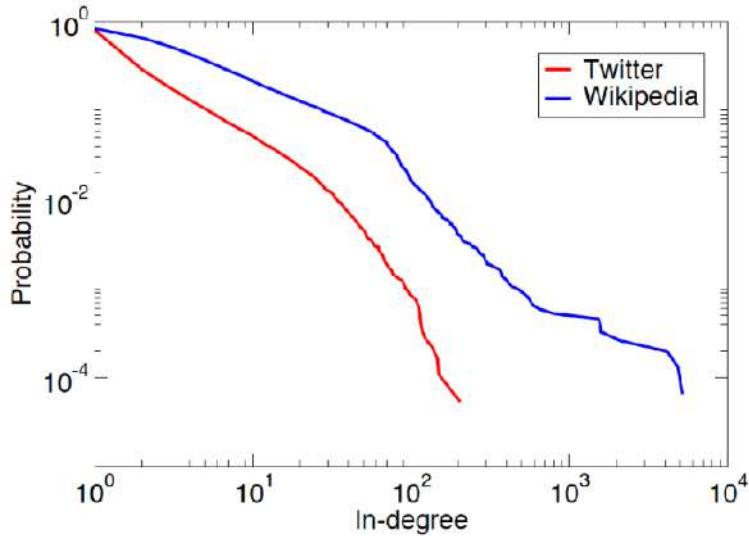
We want to understand how to plot a probability distribution if the variable spans through a large range of values starting from a very small value, and for this we can use the **logarithmic scale**. Another interesting thing is that the probability of finding larger values points in the right part of the plot is not neglectable, and this means that we will always find (if the population grows larger) the points to be plotted on the right side of the plot and we have an **Heavy Tail** distribution.

The trick is to use the logarithmic scale on the x or y axis and hence we have linear-log scales or log-log scales in our plot. The good thing about logarithmic scales is that they grow small.

$$\begin{aligned}\log_{10} 10 &= 1 \\ \log_{10} 1,000 &= \log_{10} 10^3 = 3 \\ \log_{10} 1,000,000 &= \log_{10} 10^6 = 6\end{aligned}$$

We can use log-log scales in degree distributions and have a better understanding of the graph because if we didn't we would see a curve that rapidly decreases towards the x axis and it apparently follows the x axis without doing anything.

27.3 Heterogeneity Parameter (Degree Distribution)



If we have Heavy Tail distributions (like above) it means that we have inequalities, and so we have few nodes that have high values, but the majority of the nodes is towards the beginning of the graph which means that the probability to find a small degree node is quite high, and this is heavily heterogeneous.

The **heterogeneity parameter** says how broad the distribution is:

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

$$\langle k \rangle = \frac{\sum_i k_i}{N} = \frac{2L}{N}; \quad \langle k^2 \rangle = \frac{\sum_i k_i^2}{N}$$

If most degrees have the same value, say k_0 :

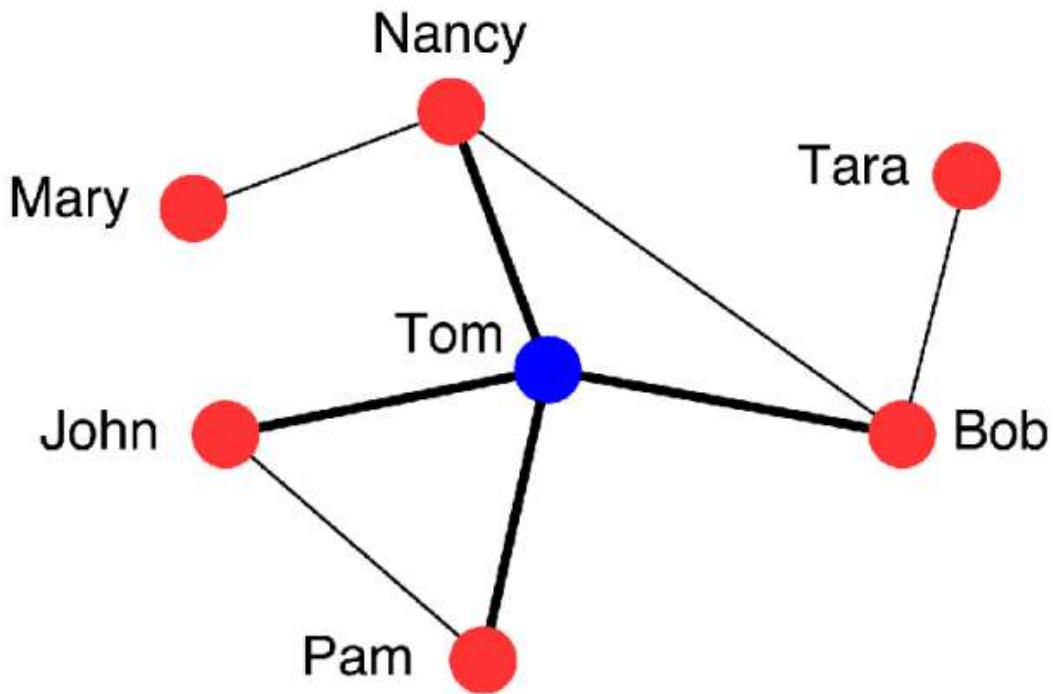
$$\langle k \rangle \approx k_0, \quad \langle k^2 \rangle \approx k_0^2 \implies \kappa \approx 1$$

If the distribution is very heterogeneous: $\kappa \gg 1$

To compute the **heterogeneity parameter** in distributions we calculate which is given by the second moment of the distributions divided by the square of the average k .

If we deal with normal distributions, the square of the k value and the second moment of the distributions have similar values and this means that if most degrees have the same value k_0 we have that the average $\langle k \rangle$ will be k_0 , the second moment $\langle k^2 \rangle$ would be k_0^2 and the heterogeneity parameter will be approx. 1. If the distributions is instead very heterogeneous the heterogeneity parameter will be much greater than 1.

28 FRIENDSHIP PARADOX



We are considering a network where a certain node is an Hub (Tom). By choosing nodes at random, inside of a network, Tom has the same chance to be picked as everybody else. But by choosing links at random, Tom has a higher chance to be picked than everybody else since he is the Hub of the network.

The idea behind the **Friendship Paradox** is that by following the links, the chance to hit an hub increases, and this has much importance in understating and controlling something like viral marketing etc.. When the network is heterogeneous I cannot assume that the randomicity that I can perform in an homogeneous network will produce the same results in an heterogeneous one. In an epidemic situation, the chance that an hub will be exposed to the virus is high and henceforth it will become a super spreader.

There is an aspect of the Friendship Paradox that makes things a less more intuitive though. If we calculate the **average degree of a node** and the **average degree of the neighbors of a node** we think that these two numbers must be the same but we see that the average degree of the neighbors of a node is much higher, and this translates to the fact that 'our friends have more friends then we do, on average' and this is the **Friendship Paradox**.

28.1 Where does the Friendship Paradox come from?

By averaging the degree of the nodes, we pick them at random. But by averaging the degree of the neighbors, we choose them by following links, which means that nodes with degree k will be counted k times, which amplifies the average. The more hubs we have the stronger the effect is, the more heterogeneous the network is, the more present the Friendship Paradox is.

29 ULTRA SMALL WORLD

Many shortest paths go through hubs and if they go through hubs that means that we have a lot of shortcuts for even very distant portions of the networks.

If we take the example of air transportation we can say:

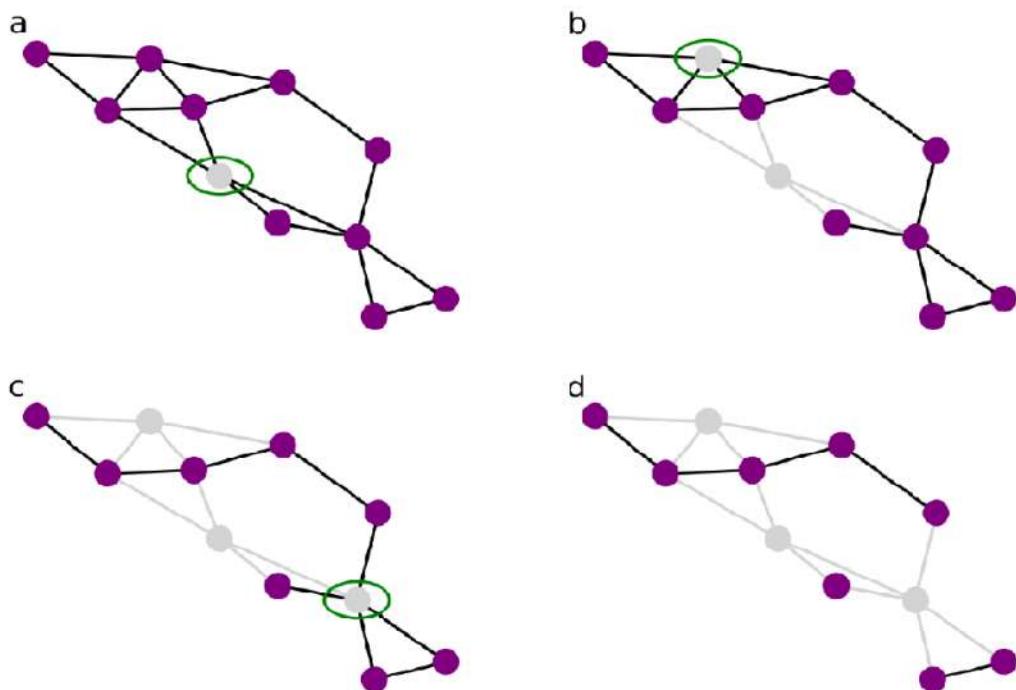
There may be no routes between airport A and B (if they are small), but it may be possible to go from A to B via a hub airport C. The small-world property is typical of most networks of interest: if the network has hubs, paths are ultra-short and this is the **ultra-small world phenomena**.

30 ROBUSTNESS

A system is considered to be **robust** if the failure of some of its components does not affect its function. This means that if we remove nodes or links in a network, the network maintains its robustness if the removal of such links and nodes won't compromise the **connectedness** of the network itself.

30.1 How can we check Robustness?

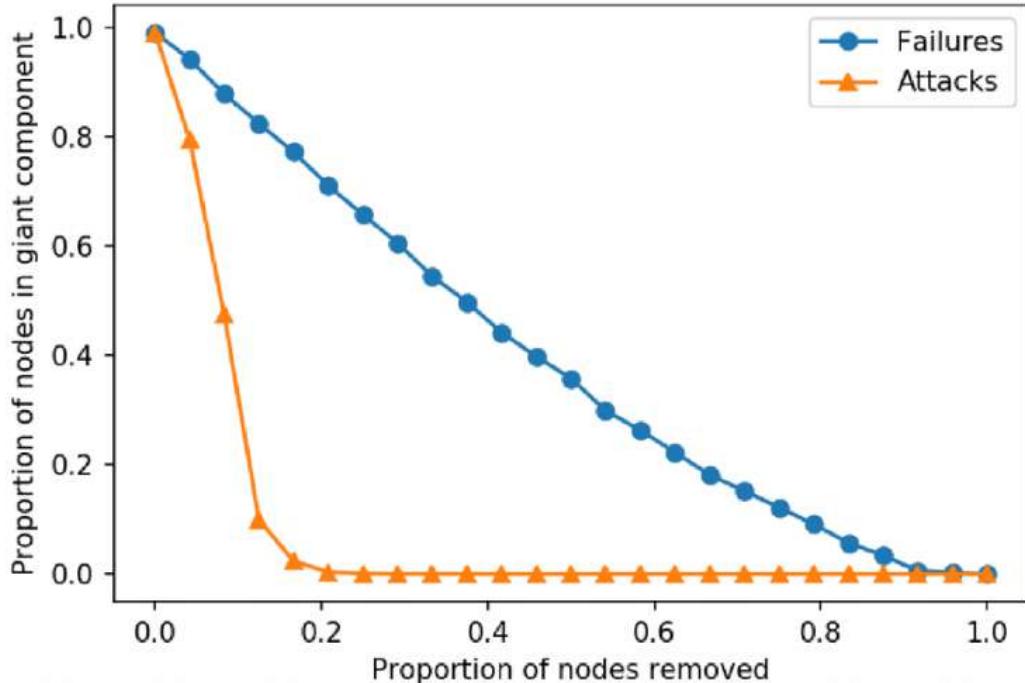
The main idea of the **robustness test** is to check how the connectedness of the network is affected as we remove more and more nodes from the network, and we want to calculate relative size of the largest connected component as a function of the set of removed nodes. If the network at the beginning is totally connected, the relative size of the largest connected component is 100 percent, but as more and more nodes and links are removed, the network starts to break up into components and the size of the largest connected component decreases. What we want to evaluate is how fast this process takes place, valuating the size of the largest connected component.



If we want to test our strategies we can compare two different types of strategies:

- **Random Failures:** where nodes break down randomly, so they are all chosen with the same probability
- **Attacks:** hubs are deliberately targeted by the degree. The higher the degree, the higher the probability of removing the node.

In the random failures approach, we remove a fraction f of nodes, chosen at random. While in the attack approach, we remove the fraction f of nodes with largest degree, from the one with largest degree downwards



If we consider the random failures we see that the shrinking of the size of the largest connected component is quite linear, while the attack leads to a more rapid decrease. We can come to the conclusion that real networks are robust against random failures but fragile against targeted attacks!

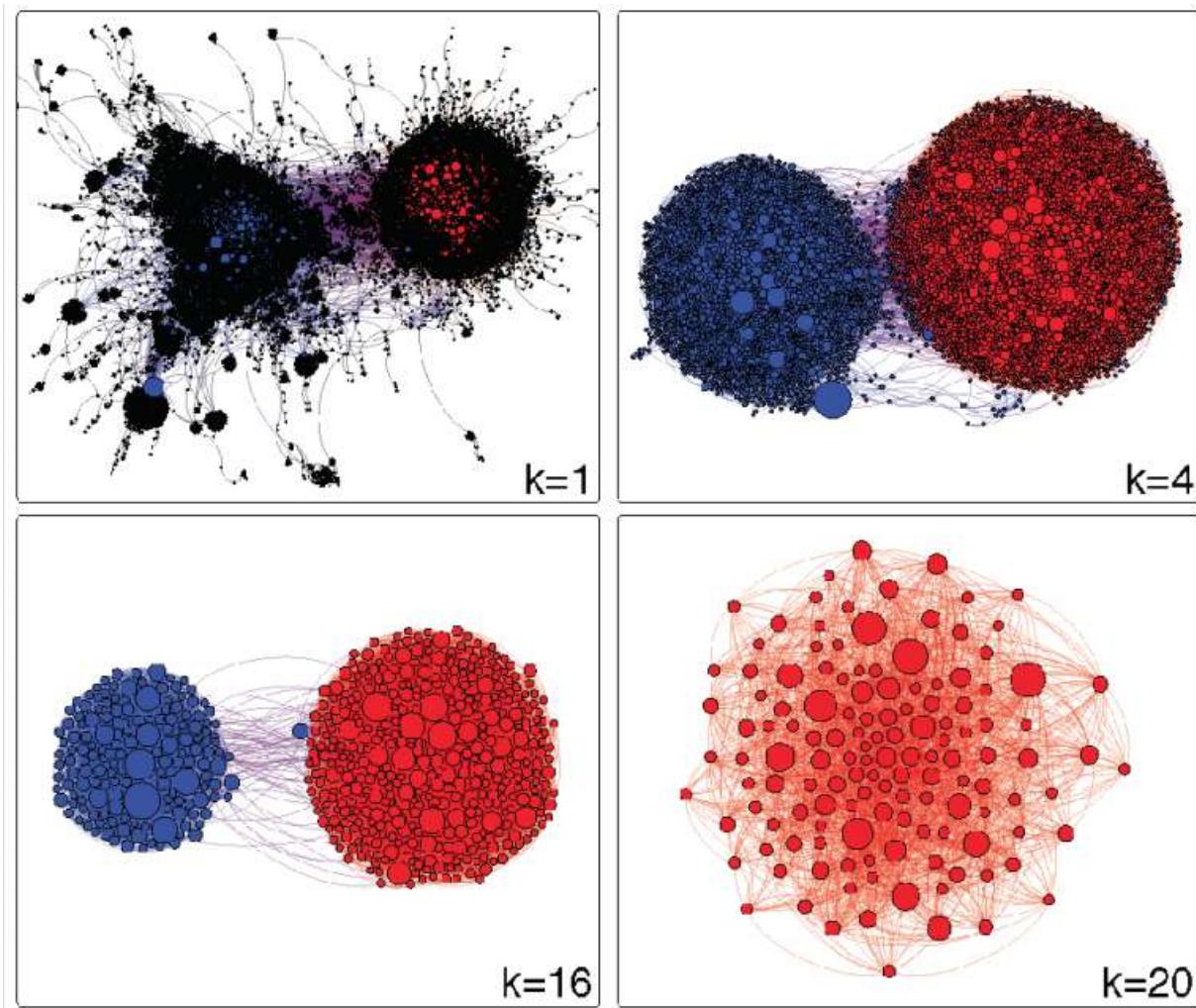
If we kill a router that has a high degree in a network, the damage we do is quite higher than removing just a random router in a network.

31 CORE DECOMPOSITION

A **Core** is dense part of the network, with high-degree nodes. **Core decomposition** is the procedure to identify denser and denser cores, by removing nodes of progressively higher degree. If we remove all nodes with degree $k-1$ or lower, the remaining portion of the network is called k -core.

The k -core decomposition procedure has three steps, starting with $k = 0$, where $k = 0$ are all the singletons:

- Recursively remove all nodes with degree k , until none are left
- The set of removed nodes is the **k -th shell**, while the remaining ones form the **$(k+1)$ -core**
- If there is no node left, terminate. Otherwise, increment k by one and repeat from step 1



32 BETWEENNESS AND GRAPH PARTITIONING

Networks are mainly made of **tightly-connected** regions connected by means of sparser connections called **Clusters**, these regions are way denser than other regions in the network, and the problem of finding these denser region is called **graph partitioning**.

We can use the structure of the network itself to predict certain fault lines, and this can be helpful to build algorithms that find clusters inside of a network and this can help me to predict crisis of future splits of my social network.

32.1 Divisive Approach

This is a top down approach where we start removing spanning links between to big components and divide a certain network in smaller parts, and go on and on until I can have a number of clusters that are as much as nodes, hence all the nodes in the network become singletons.

32.2 Agglomerative Approach

This is a bottom up approach where we form groups merging nodes with other nodes. We can start from every node and decide which approach to use, maybe we can use triadic closure to close groups and etc..

32.3 Which edge to Remove First?

In divisive approaches we need a rule to find **Spanning Edges**. We will use the notions gathered from the robustness test, and we want to remove the links in order to make the division process as fast as possible. **Local Bridges** are good candidates to be removed since the distance will be longer since it goes to infinity.

32.4 But if there are no Local Bridges?

Instead of removing local Bridges we can use the notion of **Betweenness** (B_{ij}) which is the number of shortest paths that cross through (i,j). The idea is to calculate how many shortest paths will be affected if a certain edge is removed from the network, hence betweenness is a better candidate for removal.

Edge betweennes is of course connected to the notion of flow of traffic inside of a network, if a node with high betweenness is removed, the congestion inside the network varies. High betweenness is highly correlated to **local bridges**, and to **Granovetter's weak ties theory**. It is also correlated to neighborhood overlap and to **clustering coefficient** of its endpoints. Nodes that are instead on the edge of **structural holes** are nodes that connect loosely clusters that without those nodes would be disconnected, so edges with high betweennes are probably at the border of a structural hole of a big network.

33 GIRVAN NEWMAN ALGORITHM

It is pretty straight forward:

- 1. Calculate betweenness for all the edges
- 2. Find edges with highest betweenness
- 3. Remove those edges
- 4. if satisfied or no more edges return components as clusters
- 5. else goto 1.

I could go on and remove all the links and be left with just singletons, but if we stop before we can have clusters that are useful to gather information.

We don't need to find local bridges in this approach since we use betweenness as approximation, and this is efficient since it is calculated through a global property of the network. We can stop our iterations of the algorithm basing our decision on many different critiria that we can define.

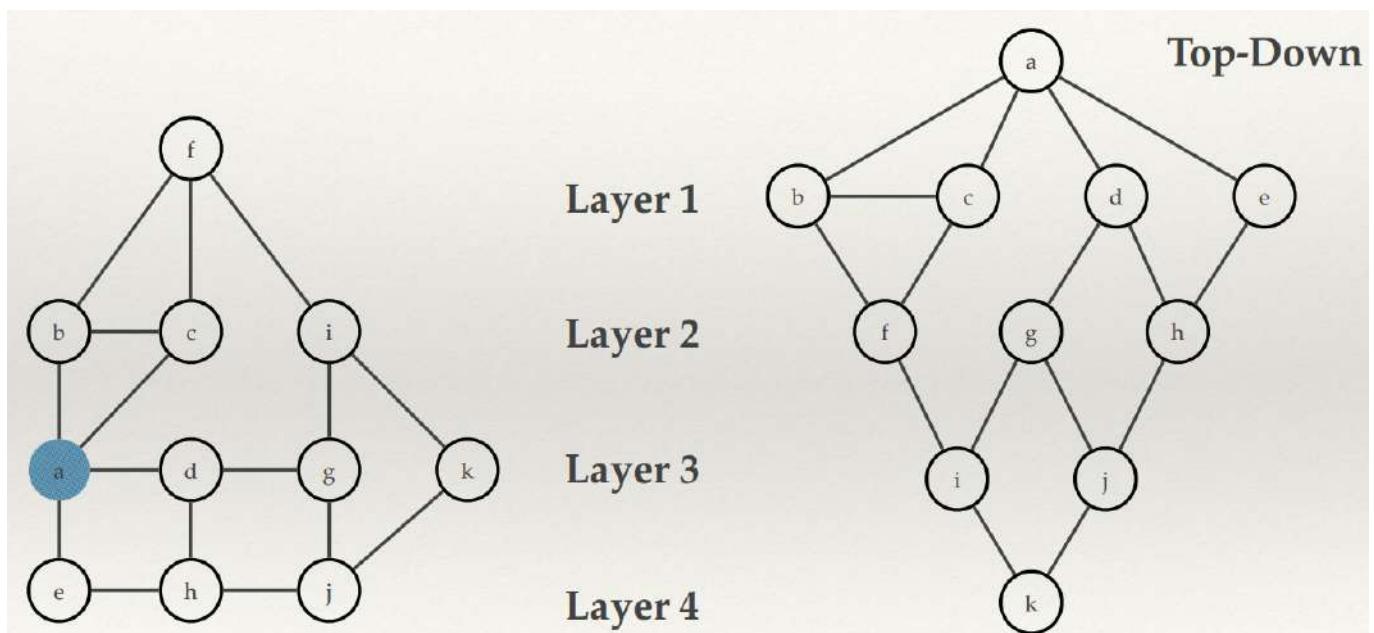
33.1 Computing Edge Betweenness Efficiently

For each node i belonging to the network:

- 1. Perform BFS (Breadth First Search)
- 2. count the number of shortest paths from i
- 3. calculate the amount of flow from i

Perform BFS (Breadth First Search):

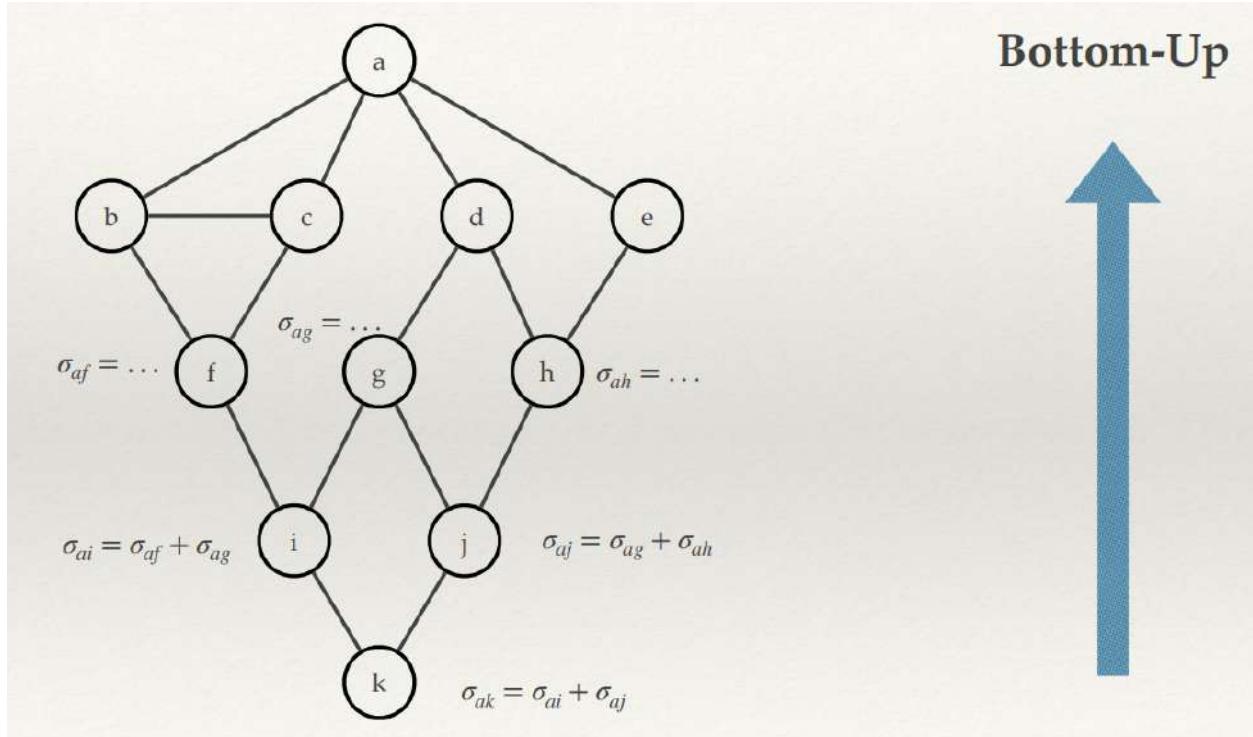
We use a top down approach starting from a specific node of the network and then we try to reach all of the nodes of the network. We get layers of nodes, and from the nodes in a certain layer we can move to other nodes that are the neighbors of the previous layer neighbors.



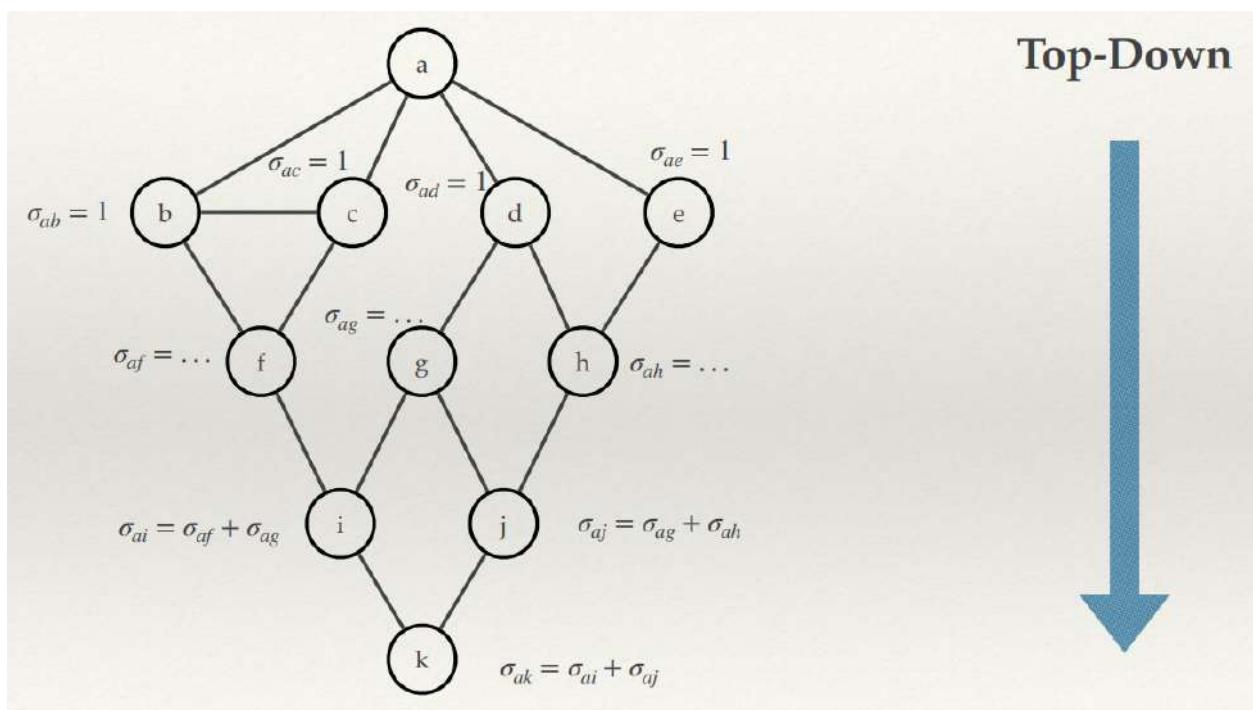
Now we go to the second step and count the number of shortest paths

Count the number of shortest paths from i:

We use in this case a bottom up approach. If we want to compute the number of shortest paths (sigma) from the starting node and a node in the last layer, the number of shortest path is just the number of the shortest path from the starting node to the nodes that precede the node in the last layer, and so on and so forth.

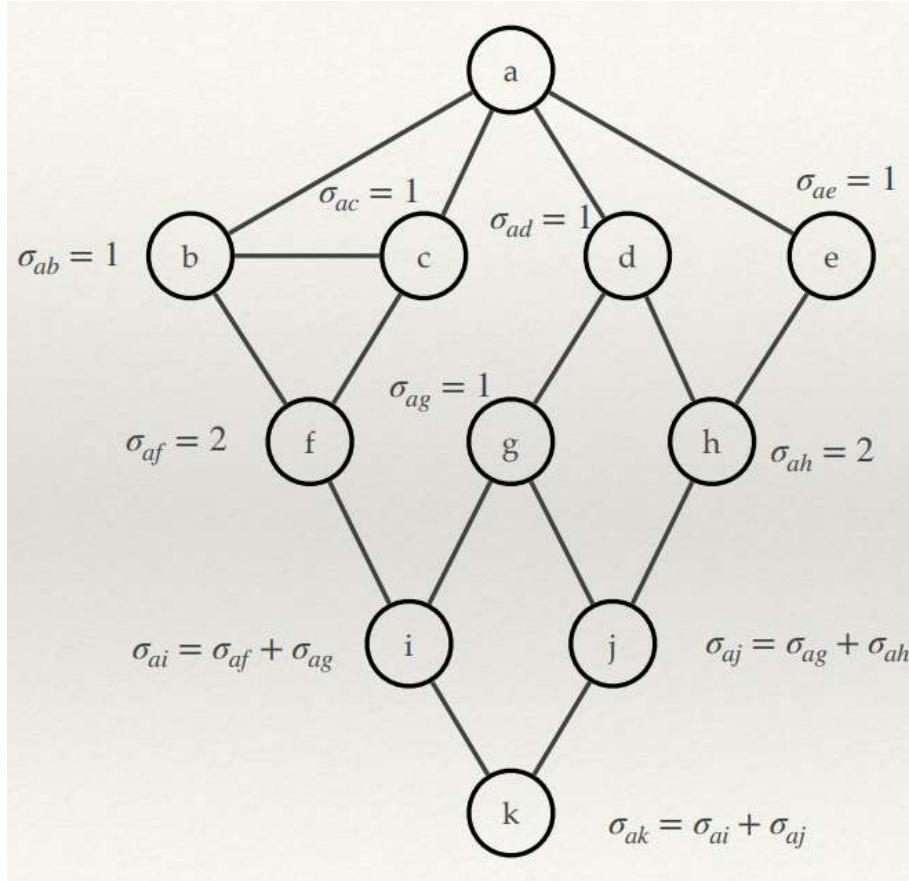


When we come at the direct neighbors of a we have only 1 shortest path and from here we can start to sum up the shortest paths in a top down method



Calculate the amount of flow from i:

The idea is that for each node i in the BFS structure, we add up all the flow arriving from edges directly below i , plus 1 for the flow destined for i itself. We then divide this up over the edges leading upward from i , in proportion to the number of shortest paths coming through each



33.2 Consideration of Girvan Newman Algorithm

If we calculate the BFS for all the nodes in G this means that I have a certain flow value for the edge (i,j) and the same flow value for the edge (j,i) , so we need to sum up all the flow values and divide them by two and get the final betweenness algorithm.

For all the nodes in the graph we need to compute a BFS that has a complexity of $O(N^2)$ so if we need to repeat the BFS for all the nodes, the complexity of this Girvan Newmann Algorithm is $O(N^3)$ for **dense graphs** and $O(N^2)$ for **sparse graphs**.

34 NETWORK MODELS

A model is a set of instructions needed to build a network. The goal is to find models that generate networks with the same characteristics as the real world networks. We need models that generate artificial networks with the same characteristics of the real networks. We start from very simple assumptions, and if we are able to predict high level global characteristics from this assumptions, we get the reasons of why this phenomena emerges in large scale networks.

We can artificially create networks, and whenever we create a network from scratch we have no reference to **real data**, and we want to understand why in certain analysis we need some artificial data to perform more precise tasks and analysis.

We need mathematical models as well, and a model is just a simplified description of a phenomena. A model can be either symbolic, categorical etc.. but in network analysis we use mathematical models.

These models are good since they assist us with computation, calculations and predictions, so with models we can deal with predictions even if we are lacking some data. Models are needed especially when we have an absence of data, like predicting a dynamic process that hasn't produced data so far like the emergence of crisis of a state or a global pandemic, we need to use models since we don't have concrete and true data yet.

It is a difficult task since we need to make predictions of something that could really happen in our world but we don't have any way to validate such data unless we just stay and wait for that data to be produced.

When we start doing predictions in a social environment it's not like the weather forecast, because social networks can react to the predictions that are being made and this is well known in economical studies (like predicting the increase of the stock market value, and this prediction itself can change the prediction of the market itself).

Models need to be used with caution, to be used always in precise environment with awareness and understanding, and we need models to predict phenomena that has already been observed.

35 FEATURES OF REAL NETWORKS

35.1 Small World Property

Most of the real networks are **small worlds**, which means that these networks show **very short average path lengths and very short distances between nodes**.

35.2 High Clustering Coefficient

$$C(i) = \frac{\tau(i)}{k_i(k_i - 1)/2} = \frac{2\tau(i)}{k_i(k_i - 1)}$$

The clustering coefficient of a node is the fraction of pairs of the node's neighbors that are connected to each other where $\tau(i)$ is the number of triangles involving i . In this definition, the clustering coefficient is undefined if $k_i < 2$, which means that a node must have at least degree 2 to have any triangles.

If we calculate the clustering coefficient of the network we can see that for social networks this coefficient is usually high, but in biological networks the coefficient is quite low, while in bipartite networks like the IMBD movie star the coefficient is usually 0. We can see heterogeneity in these clustering coefficients but the idea is that most of the times social networks have high clustering.

35.3 Heterogeneity

This is another feature of real networks, this means that if we plot the degree distribution it shows an heavy right tail which means that the number of hubs (that are nodes whose degree is higher than the ones of the other nodes in the network) is low but not neglectable. Heterogeneity can be estimated with the heterogeneity parameter (k greca) = $\langle k^2 \rangle / \langle k \rangle^2$, which is the second moment of the degree distribution divided by the square of the average degree. When the distribution is homogeneous this number is close to 1.

36 RANDOM NETWORKS

We study the Erdos Renyi model that was introduced and developed by a great mathematician Paul Erdos. The Erdos-Renyi model is the most famous and efficient random network model, and this model is a variation of the Gilber random network model.

36.1 Erdos-Renyi Algorithm

- 1. Start with N nodes and zero links
- 2. Go over all pairs of nodes, for each pair of nodes i and j , generate a random number r between 0 and
 - If $r < p$, (where p is a selected threshold that is the probability) i and j get connected
 - If $r > p$, i and j remain disconnected

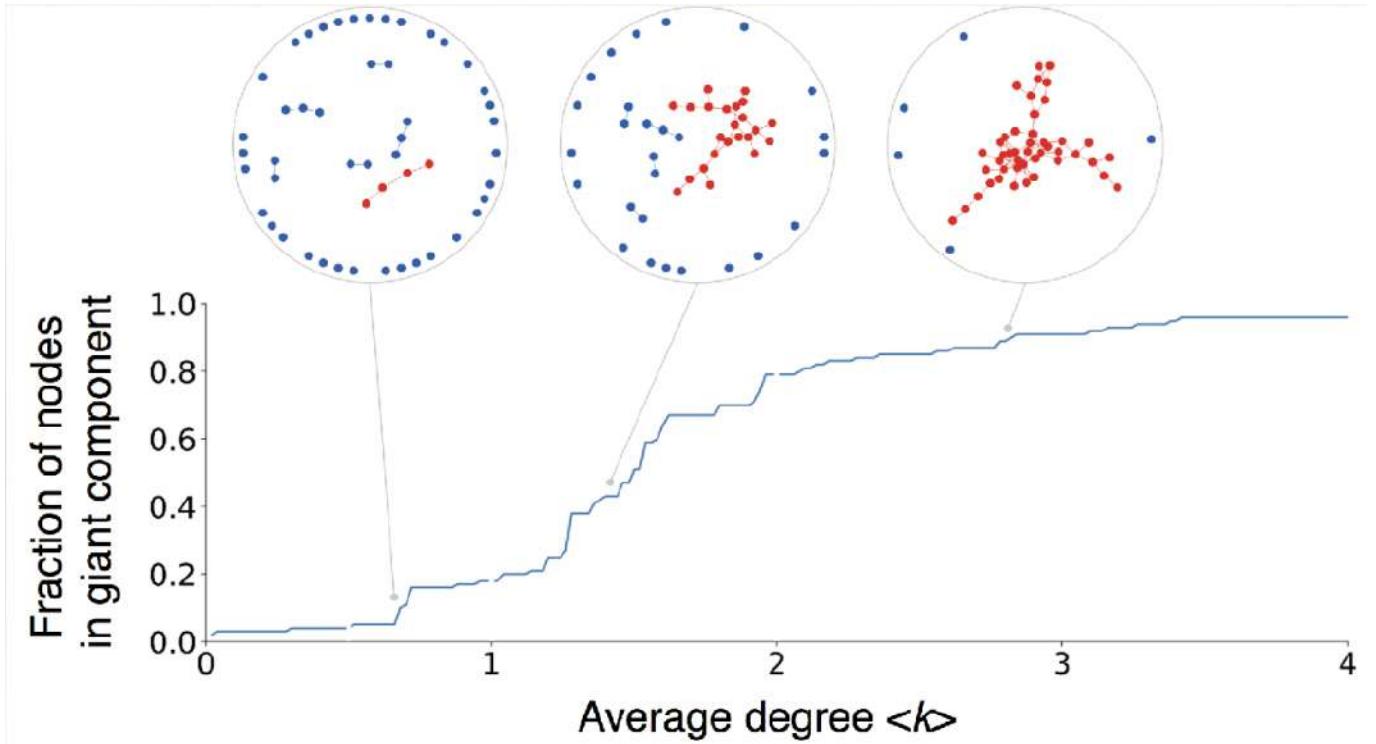
The idea behind this model is just to take a bunch of nodes, and for each pair of nodes we toss a biased coin and if we get head we establish a link, if it says tail we don't establish a link. The idea is that in this model edges are created randomly.

36.2 Evolution of Random Networks

- **If the probability p is 0**, that means that for all the pairs of nodes in the network there never is the chance to create a link, so at the end we have N components with one node each, so N singletons.
- **If the probability p is 1**, that means that every time the coin is tossed we form an edge, and hence all the links between the pairs of nodes will be created. If $p = 1$ we will have a complete network with N nodes.

As we add links to the network we can have two expectations:

- **Naive expectation:** the size of the largest component grows smoothly with the number of links
- **Wrong Expectation:** there is an abrupt increase for a given value of the link probability p



If we think about the naive expectation we expect a linear growth in the network and we see the emergence of a largest connected component, but to see that this emergence happens after the average degree surpasses 1 is counter intuitive because it is not dependent of the total number of nodes, but it depends only on the average degree. This can be generalized in networks of all sizes and orders.

37 NUMBER OF LINKS, DENSITY AND AVERAGE DEGREE IN RANDOM NETWORKS

- All the phenomena that we can extract from a random network is equivalent with a biased coin that gives heads with probability p .
- Number of independent trials (tosses) is t
- Number of heads after t trials is h
- We have some special cases where:
 - $p = 0$ means that the coin never yields heads and hence we have $h = 0$
 - $p = 1$ means that the coin always yields heads and hence we have $h = t$
 - $p = 1/2$ means that the coin yields heads half of the times and hence we have h is approximately $t/2$
- **A general rule is that h is approximately pt .**

37.1 Expected Number of Links of a Random Network with N nodes

The number of heads with probability of yielding heads equal to p and the number of trials t equal to the number of all node pairs of the network, which translates into the formula:

$$t = N(N - 1)/2 \rightarrow \langle L \rangle = p \binom{N}{2} = \frac{N(N - 1)}{2}$$

The total number of trials t is the number of possible pairs $N(N - 1) / 2$, of course in undirected network. This number is the binomial coefficient of N chooses 2 and this is the expected number of links $\langle L \rangle$ that we could have.

37.2 Expected Density of Links d of a Random Network with N nodes

$$d = \frac{\langle L \rangle}{N(N - 1)/2} = \frac{pN(N - 1)/2}{N(N - 1)/2} = p$$

The density can be calculated dividing the total number of links $\langle L \rangle$ with the total number of possible edges, and the density becomes exactly equal to the probability, and this is good because if we want to create a sparse network we need to keep the probability low, since it is exactly the density of the network. **If we want to compare an artificial network to a real network we need to keep it sparse, since real networks are mostly sparse.**

37.3 Expected Average Degree $\langle k \rangle$ of a Random Network with N nodes

Is the number of heads with probability of yielding heads equal to p and the number of trials t equal to the number of potential neighbors of a node

$$t = N - 1 \rightarrow \langle k \rangle = p(N - 1)$$

If the maximum number of trials is $N - 1$ (if we assume to be dealing with a simple graph where for every node we have one edge), this means that the average degree $\langle k \rangle$ is just $p(N-1)$.

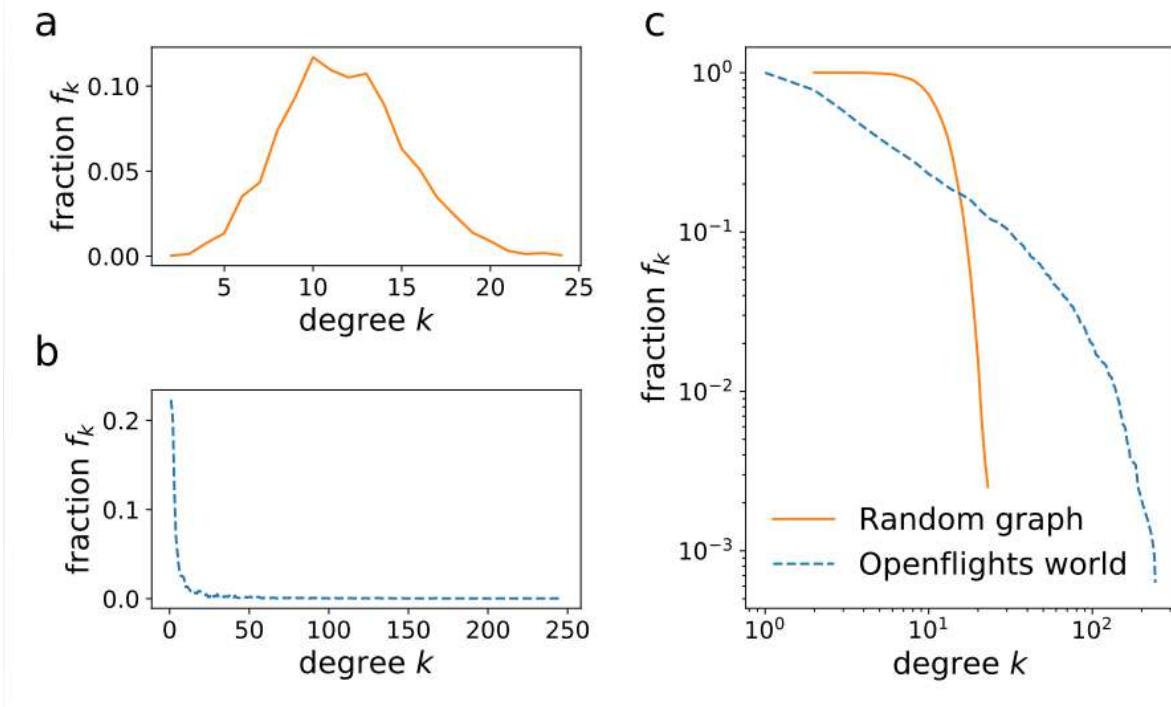
37.4 Degree Distribution in Random Networks

We want to understand what is the probability for a node to have k neighbors. If we pick up a node in a randomly generated network, we want to understand the probability that this network has a probability of k.

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

This is like having k neighbors for each nodes, and this means that we could have for k times in a group of $N-1$ different possibilities, that k out of $N-1$ would be the number of established edges. This is just the binomial coefficient of $N-1$ chooses k, and we have that with the probability p for k times (p^k) we have the creation of a link and we have the complementary probability ($1 - p$) that for the rest of the other possibilities $N - 1 - K$ we don't have the establishment of a link. **The Average Degree distribution of a Random Network is governed by a binomial distribution.**

For small p and large N the binomial distribution is well approximated by a bell-shaped curve, **this means that most degree values are concentrated around the peak, so the average degree is a good descriptor of the distribution.**



The degree distribution of random networks is very different from the broad distributions of most real world networks. The average degree for a random network is a good predictor for the population. In real worlds we have heavy tailed distributions with not many hubs, while in random graphs we have different plots where the degree follows something like a bell shaped line.

With random networks we cannot predict the heterogeneous behavior of the degree distribution.

37.5 Small World Property in Random Networks

Random Networks are good at predicting small world properties. We want to understand how many nodes are, on average, d steps away from any node in the network.

We can start from a node with a degree k , and we know that in a random network we can say that this degree k can be the average degree of the network. This means that:

- At distance $d = 1$ there are k nodes
- At distance $d = 2$ there are $k(k - 1)$ nodes
- ...
- At distance d there are $k(k - 1)^{d-1}$ nodes

If k is not too small, the **total number of nodes within a distance d from a given node** is approximately:

$$N_d \sim k(k - 1)^{d-1} \sim k^d$$

Where there is an upper bound that is $k(k - 1)^{d-1}$. At a given distance d we can find a total number of nodes that is approximately k^d .

37.6 How many steps does it take to cover the whole network?

$$\begin{aligned} N &\sim k^{d_{max}} \\ \log N &\sim d_{max} \log k \\ d_{max} &\sim \frac{\log N}{\log k} \end{aligned}$$

The total number of nodes N is approximately $k^{d_{max}}$, and if we apply the logarithmic operation we have that $\log N$ is approximately equal to d_{max} of $\log k$, which means that d_{max} is approximately equal to $\log N / \log k$ and **this means that the diameter of the network grows like the logarithm of the network size**.

37.7 Clustering Coefficient in Random Networks

The clustering coefficient (C_i) of a node i can be interpreted as the probability that two neighbors of i are connected and is the number of pairs of connected neighbors of i / number of pairs of neighbors of i .

If we want to compute what is the probability that two neighbors are connected we need to understand that since links are placed independently of each other, the clustering coefficient is just the probability p that any two nodes of the graph are connected, which is:

$$C_i = p = \frac{\langle k \rangle}{N - 1} \sim \frac{\langle k \rangle}{N}$$

Where p is the probability used in the Gilbert Model that is $\langle k \rangle / N - 1$ that is approximately $\langle k \rangle / N$. Since $\langle k \rangle$ is usually a small number, the average clustering coefficient of random networks with realistic values for $\langle k \rangle$ and N is much smaller than the ones observed in real-world network

38 SUMMARY OF RANDOM NETWORKS

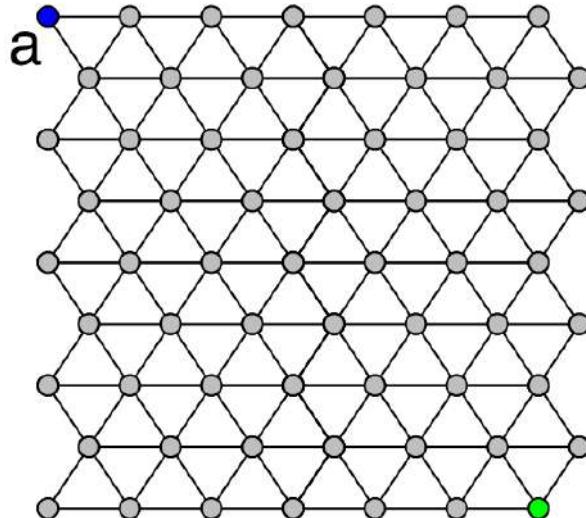
- Links are placed at random, independently of each other, and this is the simplest hypothesis we can have to build a random network
- Distances between pairs of nodes are short (small-world property) and this is a good thing
- The average clustering coefficient is much lower than on real networks of the same size and average degree and of course this is bad for the model
- The nodes have approximately the same degree, there are no hubs and this is bad since we can't predict the formation of hubs
- The conclusion is that the random networks are not a good model of many real-world networks.

So random networks are not very powerful to model a real world network since they don't represent precisely two important features of the real world networks, **but they are very useful for practical purposes**. If we need to validate the presence of a phenomena in a real network we can compute our measures and if we want to check something like checking the length of a shortest path, the best approach is to take the total number of nodes and links of the real network and create an Erdos-Reyni random network, and on the random network we compute the distances, the average clustering coefficient etc.. and we can also plot the degree distribution. If we observe that the distances in the real network are similar as in order of magnitude of the distances in the random network, we can say that the network follows the small world principle. Similarly if we compare the clustering coefficient and see that the real one is higher than the one on the random network, we can state that the clustering coefficient is high respect of the random network, and ect.. for heterogeneity.

Random networks are not toys for theoretical tasks, but they are useful to give scales about estimations we make, and to assess if some property is held or not in a network.

39 SMALL WORLD NETWORKS

Our goal is to build networks with the small world property and high clustering coefficient by shifting between a regular lattice (high clustering) and a random network (small-world property).

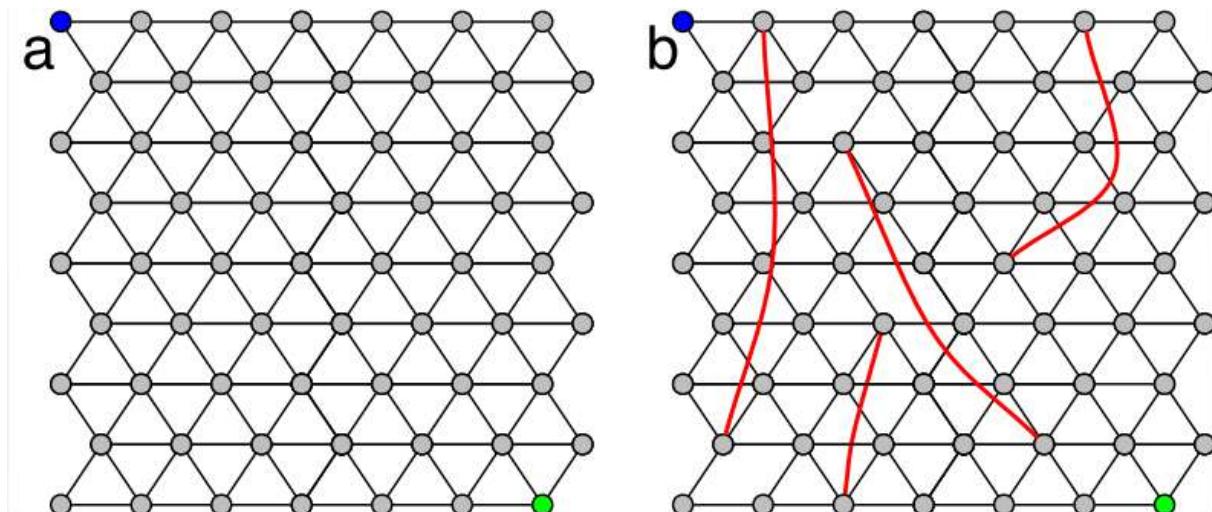


In lattice we can have peripheral nodes (a) that don't have 6 neighbors, but the majority of nodes have.

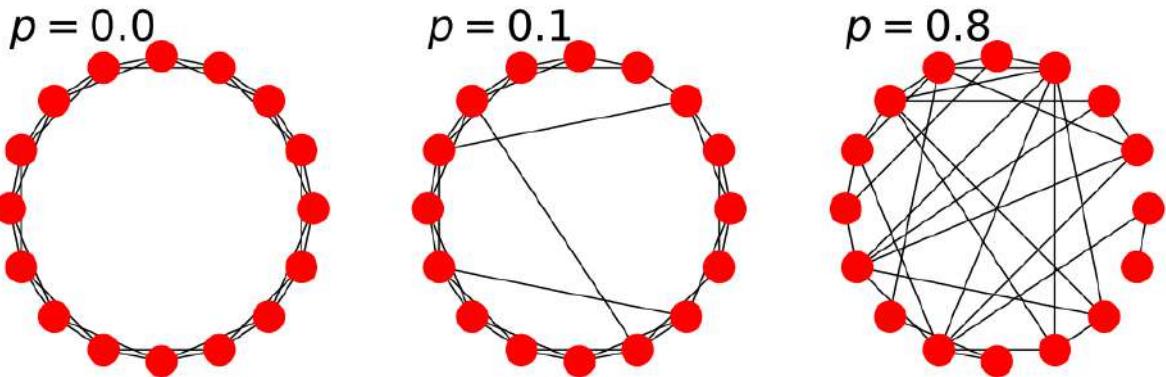
Clustering coefficient of lattice is high since the internal nodes have $k = 6$ neighbors, 6 pairs of which are connected . $C = 6/[(6*5)/2] = 6/15 = 2/5 = 0.4$ Most nodes are internal, so the average clustering coefficient of the network is close to 0.4.

Going from a node to another can take a large number of steps, which grows rapidly with the size of the network/grid and hence forth we **have a large average shortest path length**.

The idea behind small worlds is to lower the APL is to remove randomly some edges and rewire them to a randomly picked up node in the network. If we do this step we observe the creations of **shortcuts**.



39.1 The Watts-Strogatz Model



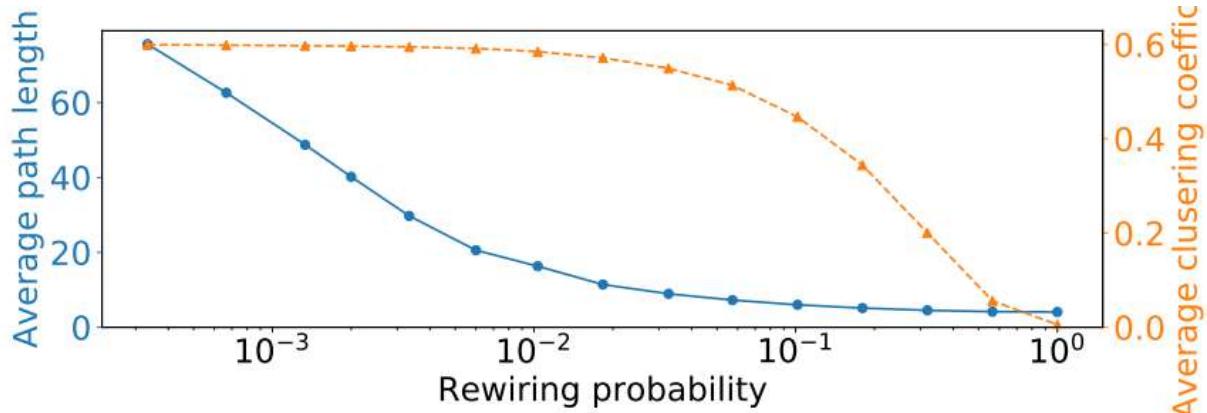
They started from a ring lattice, with N nodes that form it, with a degree k . With a probability p , the links are rewired randomly. The higher is p , the more the links are rewired, with $p = 1$ the whole edges are rewired.

The expected number of rewired links is $pL = pNk/2$, and we have that:

- If $p = 0$, no links are rewired and hence we have no change
- If p is small, few links are rewired and hence the average clustering coefficient stays approximately the same because very few triangles are destroyed, but distances shrink considerably
- If $p = 1$, all links are rewired and hence we have the opposite situation and the network becomes a random network, where we have short paths and short distances.

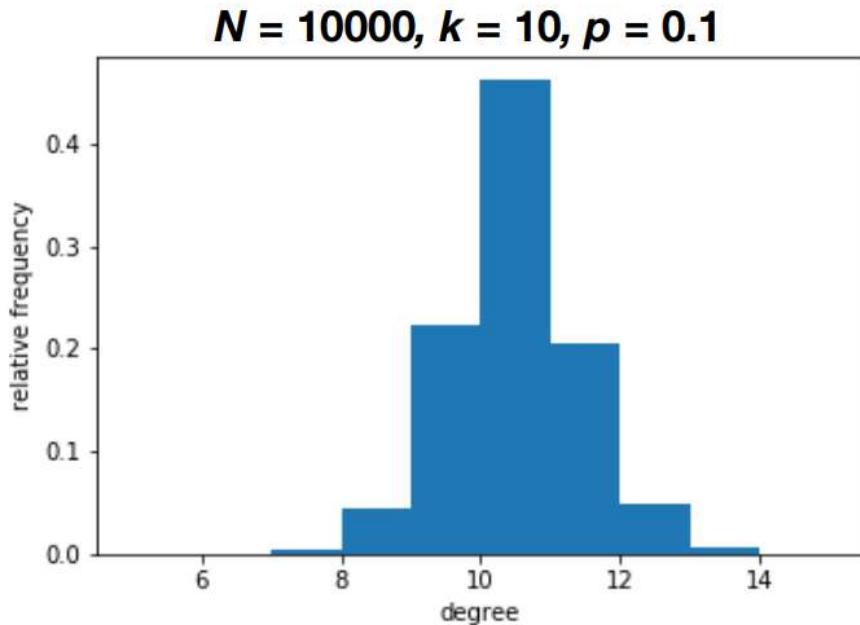
We need to find a probability p that allows us to keep an high clustering and a low average shortest path. Distances become short for low values of p , and the average clustering coefficient stays high up to large values of p .

There is a range of values of p where the average path length is short and the clustering coefficient is high.



We can hence prove that high clustering coefficient can exist in random networks where the average path length is short, because of the presence of **shortcuts**. Every node has many connections with its inner social circle, but a node can have a weak tie with a node that is further away.

39.2 Degree Distribution of the Watts-Strogatz Model



The Watts-Strogatz model fails to describe the degree distribution since this model follows a bell shaped degree distribution and we don't have heterogeneity that we see in real networks.

39.3 Summary of the Watts-Strogatz Model

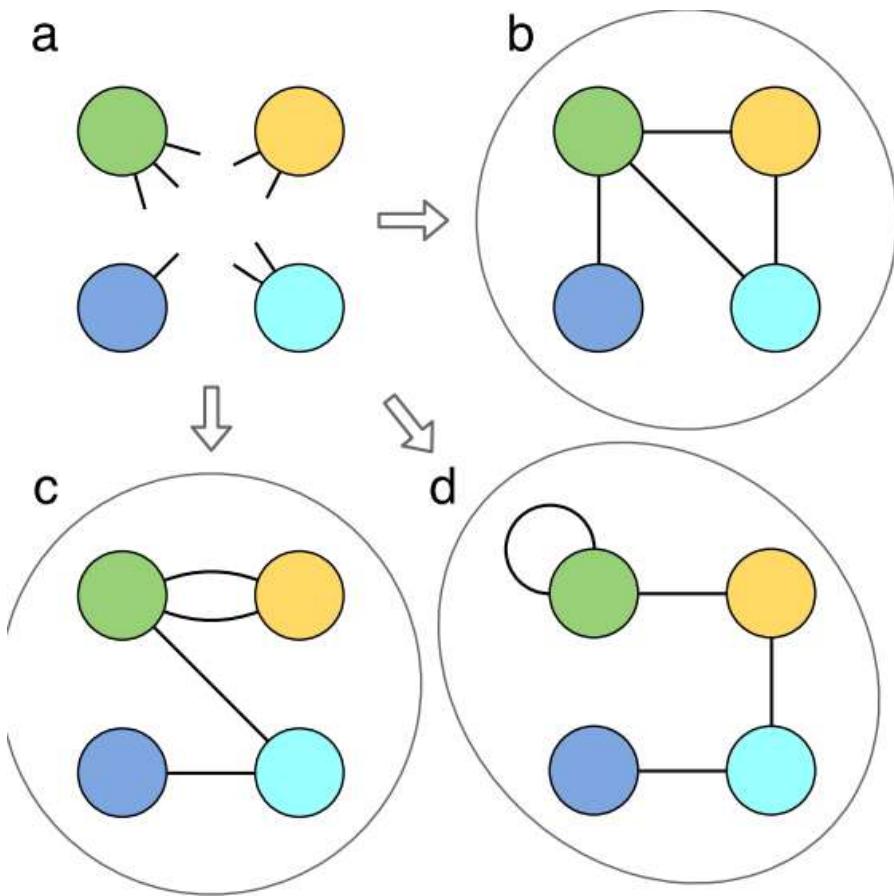
The Watts-Strogatz Model is a regular lattice whose links are randomly rewired, with some probability p .

We can conclude that:

- There is a range of values of the rewiring probability p for which distances between pairs of nodes are short (small-world property) and the average clustering coefficient is high and this is good, since it is more realistic than the Erdos-Renyi model
- But, the nodes have approximately the same degree, there are no hubs and this is quite bad

In the Watts-Strogatz Model though we have a lack of heterogeneity and to fix this we introduce the configuration model.

40 THE CONFIGURATION MODEL



The problem in this model is that we have models that are able to create networks that follow the small world principle and keep an high clustering coefficient. Sometimes though we have real networks that have a certain **degree sequence** and we want to build a random network with the same degree sequence.

The configuration model is the solution that preserves the degree distribution, since we preserve the degree for each node, but we interrupt some links. It's like we have nodes that are hubs and have links that are ready to be established, and if we take randomly pairs of nodes and try to connect them we get different realizations of networks while preserving the degree.

A degree-preserving randomization generates randomized versions of a given network with the same degree sequence, using the configuration model.

This is useful to see whether a specific property of the original network is determined by its degree distribution alone:

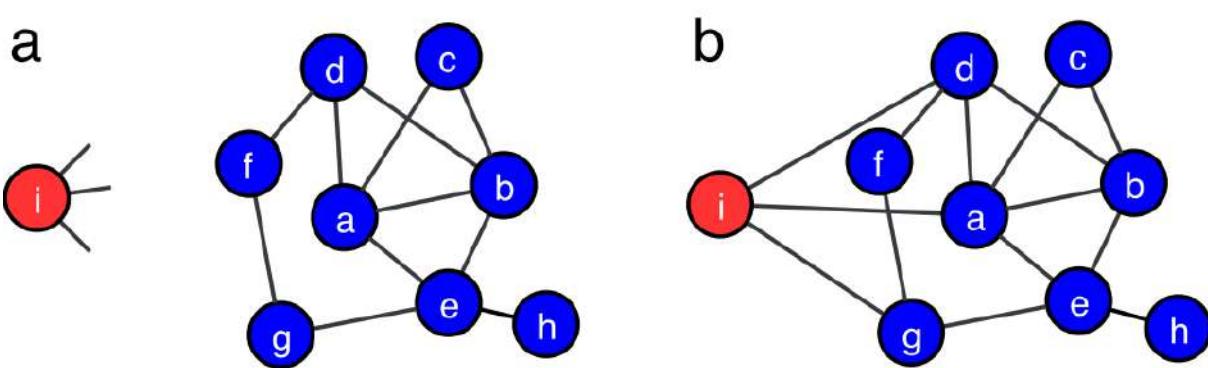
- If the property is maintained in the randomized configurations, then the degree distribution is the main driver
- If the property is lost in the randomized configurations, other factors must be responsible for it

41 PREFERENTIAL MODELS

41.1 Network Growth

Real-world networks are dynamic, and they keep growing as time flows by. The general procedure to simulate the growth of the network is:

- 1. To have a new node come with a given number of stubs (which are links not yet established), indicating the number of future neighbors of the node (degree)
- 2. The stubs are attached to some of the old nodes, according to some rule that we can specify



41.2 Preferential Attachment

This rule says that nodes prefer to link to the more connected nodes. An example is that our knowledge of the Web is biased towards popular pages, which are highly linked, so it is more likely that our website points to highly linked Web sites

The network model to be used should have two main features:

- **Growth:** the number of nodes grows in time following the addition of new nodes. The models considered so far are static
- **Preferential Attachment:** new nodes tend to be connected to the more connected nodes. The models considered so far set links among pairs of random nodes, regardless of their degree

From this we can justify the famous phrase *the rich get richer and the poor get poorer*.

41.3 Polya's Urn Model

The idea behind this model is to start with an urn contains X white and Y black balls and the process is that a ball is drawn from the urn and put back in with another ball of the same color

An example could be that if we first pick a white ball, there will be X+1 white and Y black balls in the urn; white will become more likely to be picked than black in the future

Preferential attachment is used to explain heavy-tail distributions of many quantities like the number of species per genus of flowering plants, the number of (distinct) words in a text, the populations of cities, individual wealth, scientific production, citation statistics, firm size, etc.

41.4 The Barabasi-Albert Model

We start with a group of m_0 nodes, usually fully connected (clique), and at each step a new node i is added to the system, and sets m links with some of the older nodes ($m \leq m_0$).

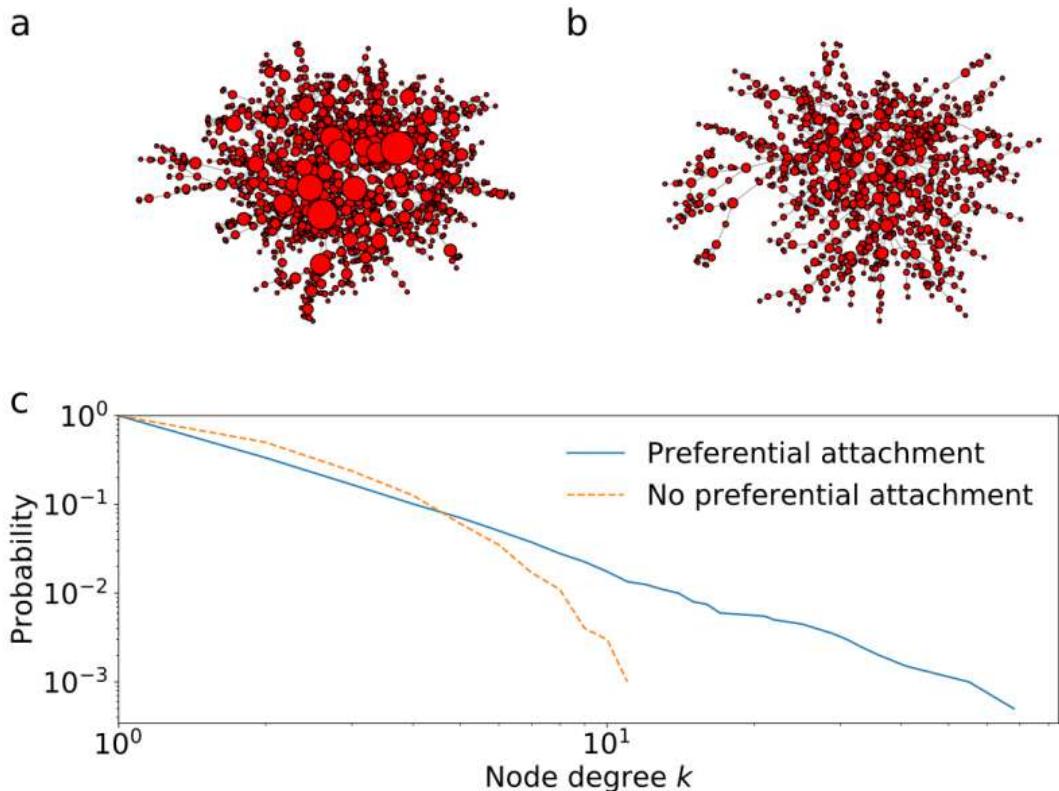
The probability that the new node i chooses an older node j as neighbor is proportional to the degree k_j of j and hence we have that if a node j has an higher degree, it has an higher probability to be selected as an endpoint:

$$\Pi(i \leftrightarrow j) = \frac{k_j}{\sum_l k_l}$$

Here we can see the rich-gets-richer phenomenon: due to preferential attachment, the more connected nodes have higher chances to acquire new links, which gives them a bigger and bigger advantage over the other nodes in the future, and this is how hubs are generated.

Hubs are the oldest nodes since they get the initial links since the very beginning of the network, and acquire an advantage over the other nodes, which increases via preferential attachment.

If old nodes have an advantage over newer nodes anyway, do we even need preferential attachment at all? Can we explain the existence of hubs just because of growth?



We can create an alternative model where each new node chooses its neighbors at random, not with probability proportional to their degree, and what happens is that if we compare two network growths, one with preferential attachment and one without, we observe that growth and random attachment don't generate hubs, and hence **preferential attachment is necessary**, we don't have hubs when we don't apply preferential attachment.

41.5 Other Preferential Models

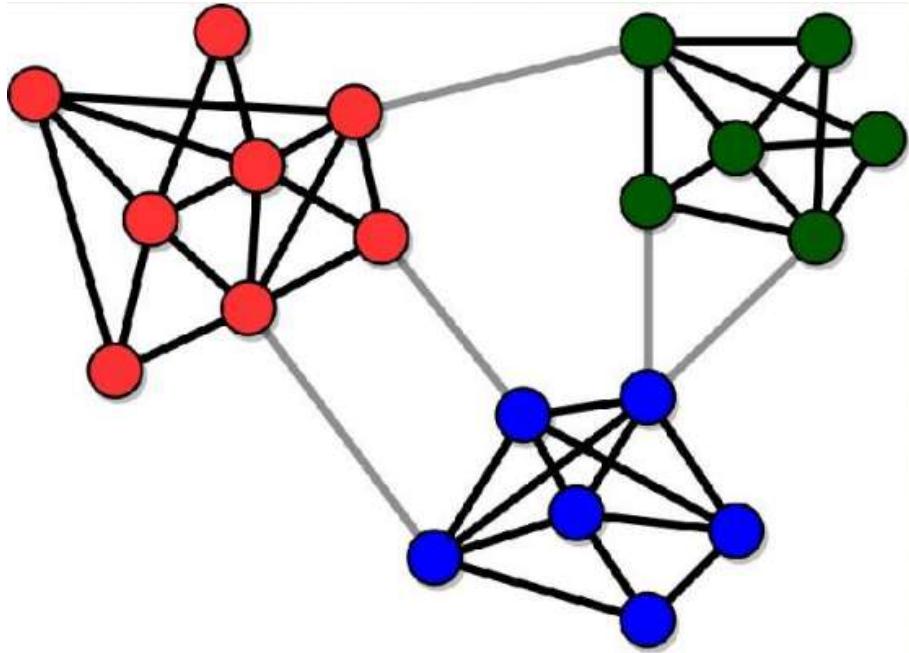
The Barabasi-Albert model uses linear preferential attachment which means that the linking probability is proportional to the degree. But what happens if the linking probability is proportional to a power of the degree? We get a totally new family of models that are called **Non-linear Preferential Attachment models**.

42 FINAL CONCLUSIONS ABOUT MODELS

- **Watts-Strogatz model** explains short paths high clustering, but fails to get how individuals actually find short paths from only a local perspective, and to solve this we can use Kleinberg's decentralized search
- **Barabasi-Albert model** generates scale-free networks, that is, degree distribution following power law with exponent $2 \leq \alpha < 3$ which is the analytical analysis of rich get richer process

43 COMMUNITIES IN NETWORKS

43.1 Community Structure



Communities, or clusters, are a set of tightly connected nodes. Finding communities is not easy and this is the main problem of this definition, because it is too weak since we can't specify so loosely what tightly connected means.

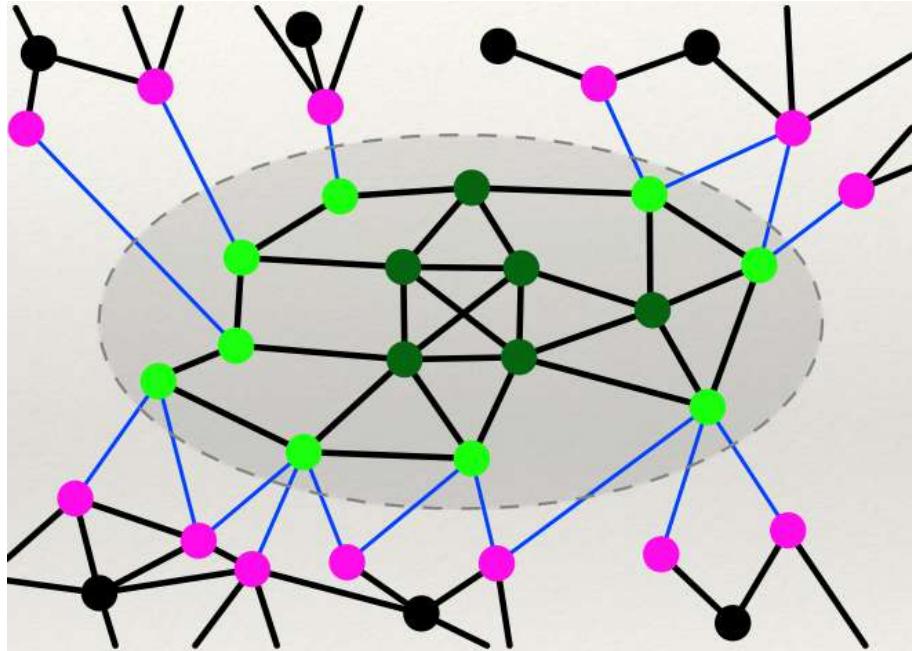
As an example of community we can specify that Twitter users with strong political preferences tend to follow those aligned with them and not to follow users with different political orientation.

43.2 Why to Study Communities?

There are some main reasons:

- Uncover the inner organization of the network
- Identify some specific features of the nodes
- Classify the nodes based on their position in the clusters
- Find missing links that should be established but that aren't and hence get conclusions from the absence of that link

43.3 Definition of Variables



- **Internal degree** of a node is the number of neighbors of the node in its community
- **External degree** of a node is the number of neighbors of the node outside of its community
- **Community degree** is the sum of the degrees of the nodes in the community
- **Internal link density** is the ratio between the number of links L_c inside a community C and the maximal possible number of links that can lie inside C:

$$\delta_C^{int} = \frac{L_c}{L_c^{max}} = \frac{L_c}{\binom{N_c}{2}} = \frac{2L_c}{N_c(N_c - 1)}$$

Where N_c is the number of nodes in C

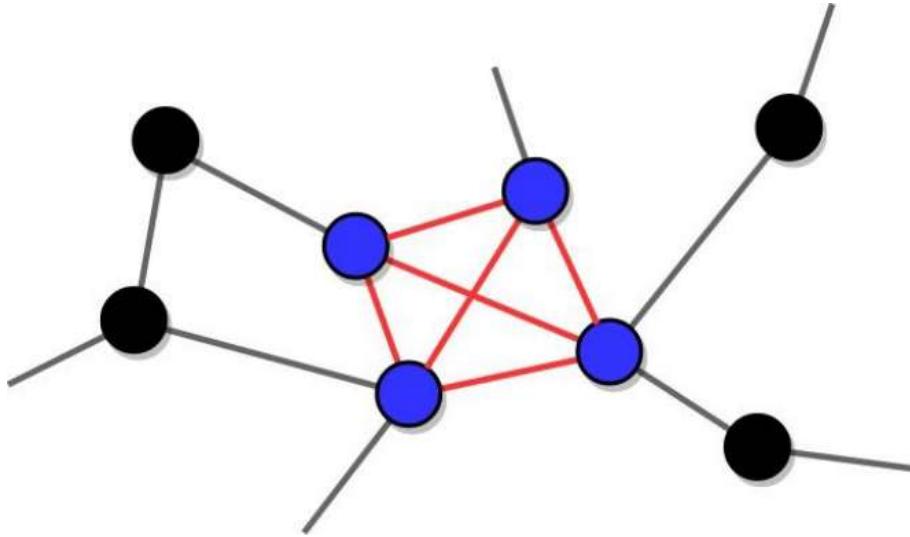
43.4 Definitions of Community

We have two main features in communities

- **High cohesion communities** have many internal links, so their nodes stick together
- **High separation communities** are connected to each other by few links

We need to have something more quantitative to build a more efficient strategy to look for communities in a quantitative way.

43.5 Community Definitions Based on Cohesion



The idea is to focus on the cluster's properties disregarding the rest of the network. As an example we can use a clique, where all of its internal links are established and we have a **maximal cohesion**.

The problem with this is that nodes are connected to all other nodes in the cluster, but in real communities they have different roles, which is reflected in heterogeneous linking patterns.

43.6 Community Definitions Based on Cohesion vs Separation

The idea is to define communities in terms of internal high cohesion and external high separation. This is done by counting the number of internal links of a subgraph and if this number exceeds the number of external links we have found a community.

We can have:

- **Strong communities** when subnetwork such that the internal degree of each node is greater than its external degree
- **Weak communities** when subnetwork such that the sum of the internal degrees of its nodes is greater than the sum of their external degrees

From these definitions we can observe that:

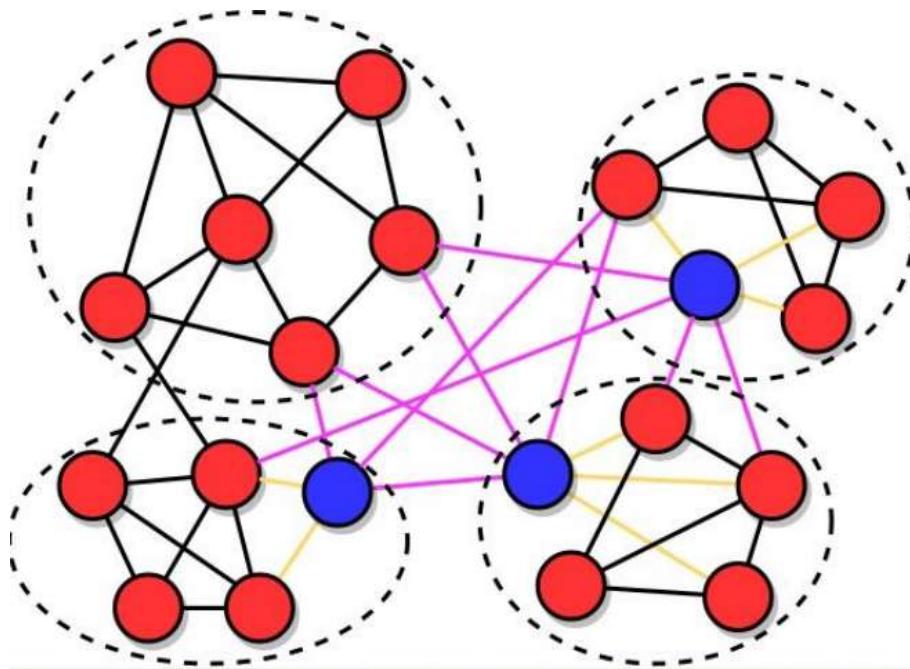
- **A strong community is also a weak community** since if the inequality between internal and external degree holds for each node, then it must hold for the sum over all nodes
- **A weak community is not always a strong community** because if the inequality between internal and external degree holds for the sum, it may be violated for one or more nodes

The issue with this definition of strong and weak communities we can compare a subnetwork with the rest of the network, but it makes more sense to compare subnetworks to other subnetworks and not the rest of the network

43.7 Better Definition of Communities

We can then give a better definition of communities where:

- **Strong communities** are subnetworks such that each node has more neighbors inside it than in any other community
- **Weak communities** are subnetworks such that the sum of the internal degrees of its nodes exceeds the total number of neighbors that the nodes have in any other community

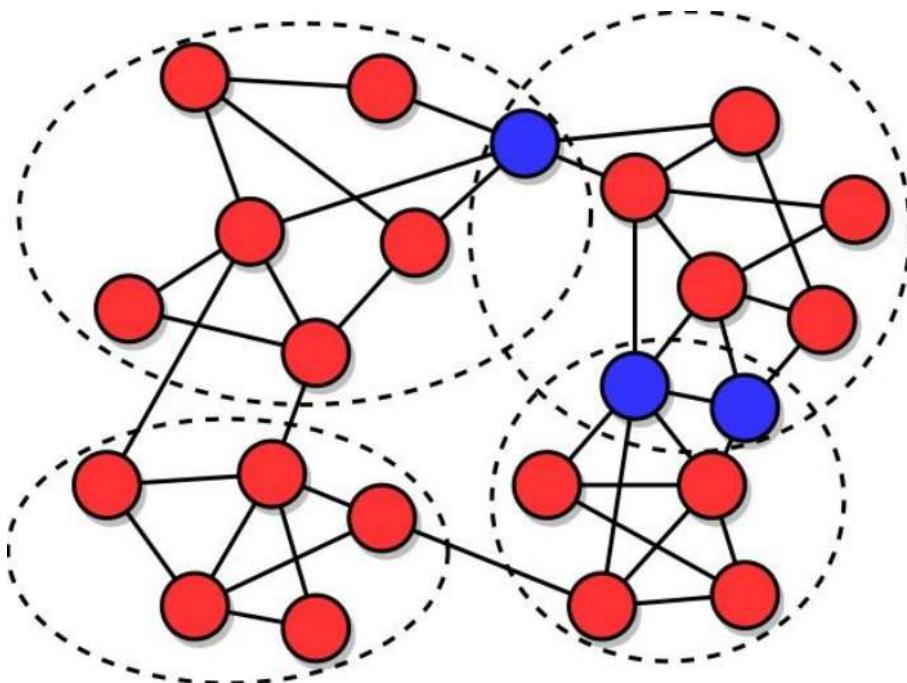


44 PARTITIONS

The simplest way to find a partition in a network is to look extensively in the space of every partitions, and the number of partitions in a graph is known to follow the Bell number. The Bell number grows more than the exponential with n , and this means that after just $n = 15$ we can have a massive number of partitions, and hence we cannot explore all the partitions but just a small subset of all the possible partitions.

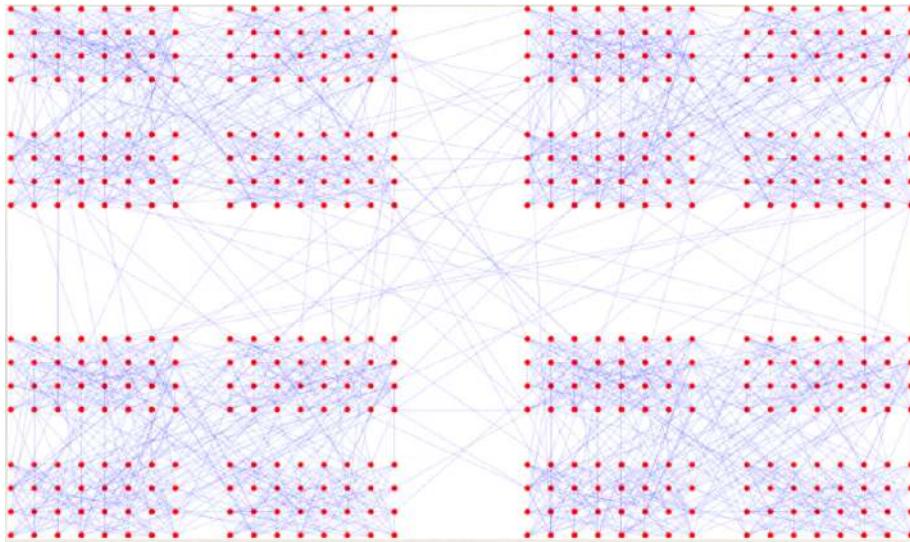
The conclusion is that it makes no sense to look for interesting community structures by exploring the whole space of partitions , just a smart exploration of the partition space must be performed.

44.1 Overlapping Communities



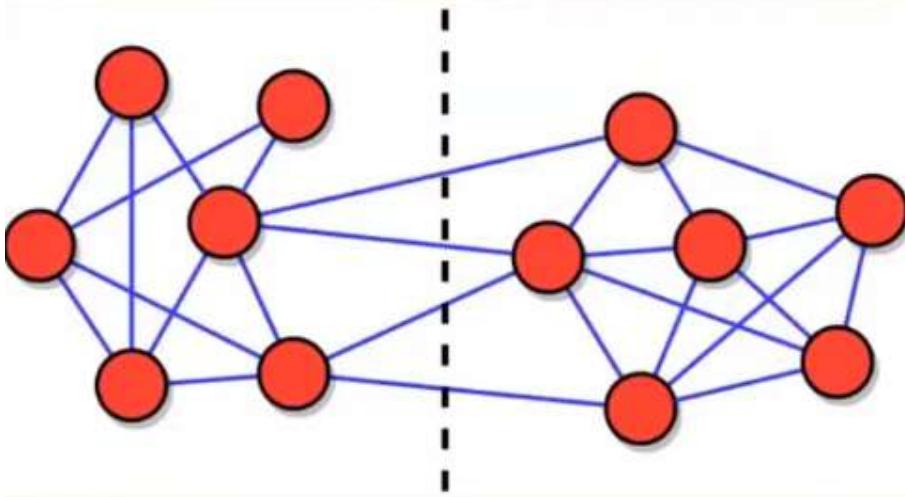
Communities in many real networks can **overlap**. A division of a network into overlapping communities is called cover. The number of possible covers of a network is far higher than the number of partitions, due to the fact that we have many ways in which clusters can overlap.

44.2 Hierarchical Communities



If the network has multiple levels of organization, its communities could form a **hierarchy**, with small communities within larger ones. An example of this could be the branches of the police organization that is divided in different sections and departments. **All hierarchical partitions are meaningful and a good clustering algorithm should detect all of them.**

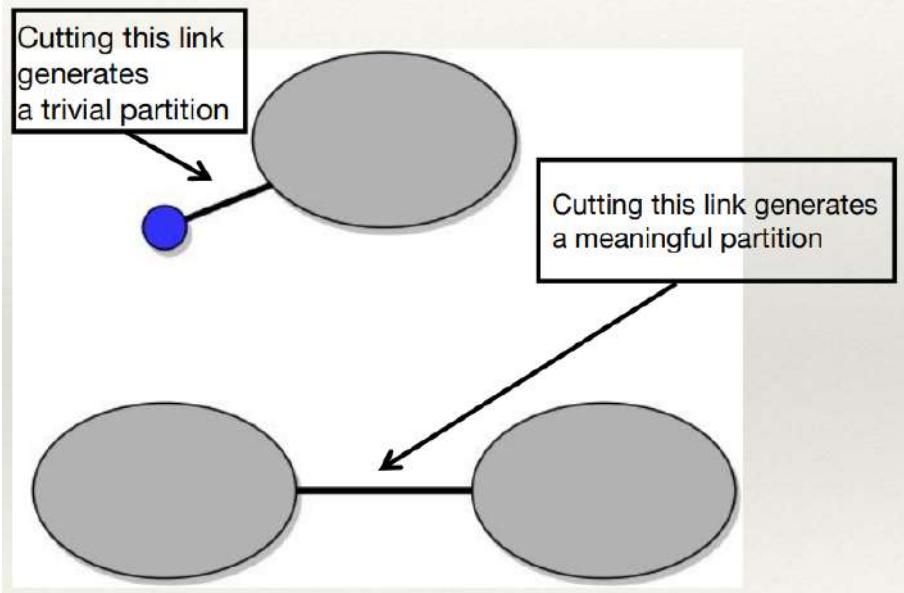
44.3 Network Partitioning



In a simplified version of community detection, we divide the nodes of a network into a number of groups of predefined size in a way that the number of links between the groups is minimal. When we have a number of nodes that compose the network, we want to find the minimum number of links that need to be removed in order to have a number of separated partitions in our network. The problem is dividing the nodes of a network into a number of groups of predefined size, such that the number of links between the groups is minimal. The number of links between the groups is called **cut size**.

44.4 Related Problems in Network Partitioning

If the number of clusters is not given beforehand, the trivial solution is to have a single cluster including everything, so the whole graph is a cluster. If the size of the clusters is not indicated in advance, there may be trivial solutions by **removing the nodes with lowest degree**.



Cutting meaningful links gives us partitions of similar sizes.

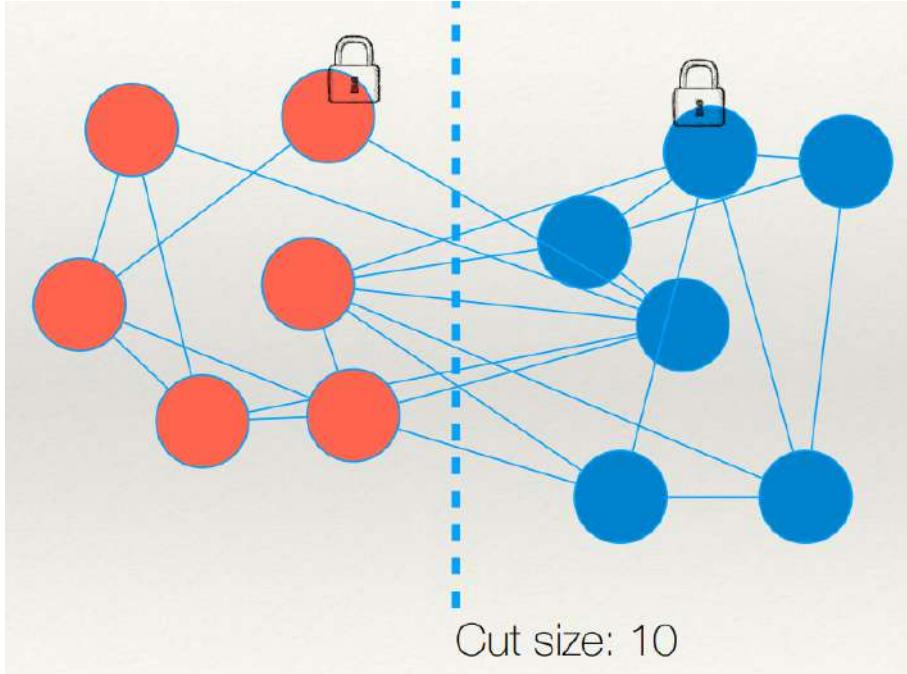
44.5 Network Bisection

The problem is that we want to divide the nodes of a network into two groups of equal size, such that the number of links between the groups is minimal.

45 Kernighan-Lin Algorithm

This is the most famous algorithm for network partitioning. And it has 4 steps:

- 1. Split in two groups A and B of predefined sizes NA and NB , by randomly assigning nodes to either group (for bisection $NA = NB$)
- 2. For each pair of nodes i, j , with i belonging A and j belonging B, compute the variation in cut size between the current partition and the one obtained by swapping i and j
- 3. The pair of nodes i and j that give the largest decrease in cut size is selected and swapped. This pair of nodes is locked, they will not be touched again during this iteration
- 4. Repeat steps 2 and 3 until no more swaps of unlocked nodes give a decrease in cut size. This creates a new bipartition, that is used as a starting configuration for the next iteration



The procedure reaches **convergence** and ends when the cut size of partitions obtained after consecutive iterations is the same, meaning that the algorithm is unable to improve the result. The algorithm can be applied to minimize the cut size of partitions into more than two clusters, by swapping nodes between pairs of clusters.

We have a problem with this approach, that is that **the choice of the initial partition heavily affects the final result**. The larger the cut size of the initial partition, the worse the final solution and the longer the time to reach convergence. **A feasible solution is to create multiple random partitions and pick the ones with the lowest cut size as initial partitions.**

This algorithm is greedy, in that at each step one looks for the partition with the smallest cut size. Because of this, the algorithm gets stuck in **local optima**, which means that we can have some solutions whose cut size is not as low as it can be. The problem can be mitigated by occasionally allowing swaps of nodes that increase the cut size.

This algorithm is widely applied as a post-processing technique, to improve partitions delivered by other methods. Such partitions can be used as starting points for the method, which might return solutions with lower cut size.

45.1 Limits of Network Partitioning

Clusters have to be well-separated, but they do not need to have high internal link density, which actually means that clusters found via graph partitioning are not in general communities.

46 DATA CLUSTERING

The idea behind data clustering is to group objects based on how similar they are to each other. We use some similarity measures and then we group the objects that have higher similarity, and of course higher the similarity, higher is the chance to link two similar nodes. We have two main classes of data clustering algorithms: **hierarchical clustering** and **partitional clustering**.

47 HIERARCHICAL CLUSTERING

This procedure delivers a nested set of partitions. For this to work we just need to define a **similarity measure**. For example, in a social network it could indicate how close the profiles of two people are based on their interests like music, films etc.. If nodes are embedded in space (which means that they are points in a metric space), the **dissimilarity** between two nodes can be expressed by their distance, and if they are not embedded in space, the similarity measure can be derived from the network structure, and hence we need a **structural equivalence**.

47.1 Structural Equivalence

The idea is that nodes are similar if their neighbors are similar.

- **Concept:** nodes are similar if their neighbors are similar

$$S_{ij}^{SE} = \frac{\text{number of neighbors shared by } i \text{ and } j}{\text{total number of nodes neighboring only } i, \text{ only } j, \text{ or both}}$$

- **Examples:**

- If the neighbors of i and j are (v_1, v_2, v_3) and (v_1, v_2, v_4, v_5) , respectively, $S_{ij} = 2/5 = 0.4$, because there are two common neighbors (v_1 and v_2) out of five distinct neighbors in total (v_1, v_2, v_3, v_4, v_5)
- If i and j have no neighbors in common, $S_{ij} = 0$
- If i and j have the same neighbors, $S_{ij} = 1$

If we have nodes i and j , we can have the number of neighbors that i and j have in common as the numerator, and at the denominator we can have the total number of nodes that only i has as neighbor, only j or even both of them. If on the denominator we have the total number of nodes that neighbor i and j , we have the **Jaccard Similarity**.

47.2 Similarity of Nodes Groups

We want to understand if we can define the similarity between two groups of nodes $g1$ and $g2$ via the similarity S between two pair of nodes. We can do this with multiple approaches. The first step is to compute the pairwise similarity S_{ij} between each node i in $G1$ and each node j in $G2$. Then the following strategies can be adopted:

- **Single linkage:** take the maximum pairwise similarity $\rightarrow S_{G_1 G_2} = \max_{i,j} S_{ij}$
- **Complete linkage:** take the minimum pairwise similarity $\rightarrow S_{G_1 G_2} = \min_{i,j} S_{ij}$
- **Average linkage:** take the average pairwise similarity $\rightarrow S_{G_1 G_2} = \langle S_{ij} \rangle_{i,j}$

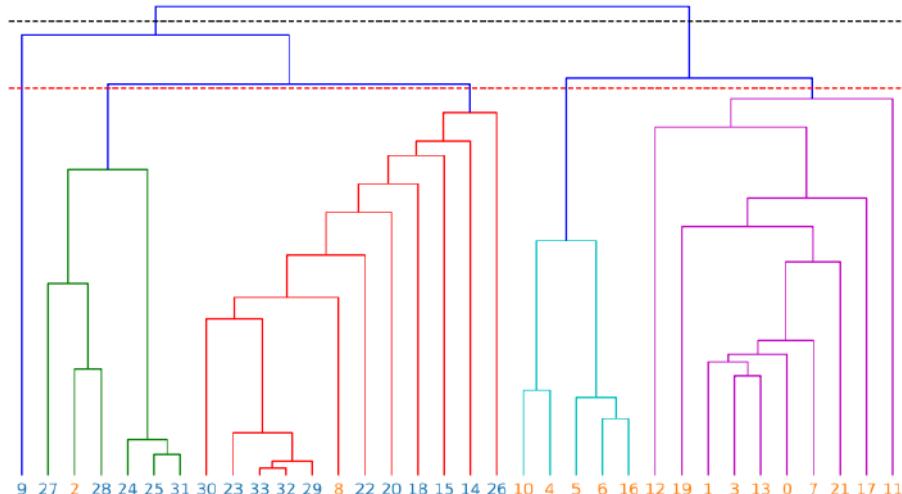
47.3 Types of Hierarchical Clustering

We have two main approaches:

- **Agglomerative hierarchical clustering:** partitions are generated by iteratively merging groups of nodes
 - Start: partition into N groups, each group consists of one node, of singletons
 - At each step the pair of groups with the largest similarity are merged , where the similarity is defined on the characteristic we want to use
- **Divisive hierarchical clustering:** partitions are generated by iteratively splitting groups of nodes

47.4 Dendrograms

The result of agglomerative hierarchical clustering gives us a dendrogram, which is a compact summary of all the partitions that are created from an hierarchical clustering. **Since each merge reduces the number of groups by one, the total number of resulting partitions is N.**



Dendrograms have certain feature as:

- At the bottom we can find the leaves of the tree, indicated by the labels of the nodes
- As we start going upwards, pairs of clusters are merged.
- The nodes of each cluster can be identified by following the vertical line representing the cluster all the way down
- Partitions are selected via horizontal cuts of the dendrogram: the clusters are the ones corresponding to the vertical lines severed by the cut
- High cuts yield partitions into a few large clusters, low cuts yield partitions into many small clusters
- Hierarchy: each partition has clusters including clusters of all partitions lying lower in the dendrogram

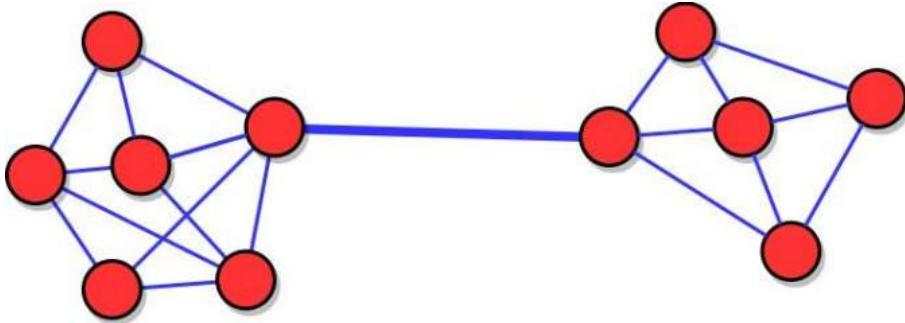
47.5 Limits of Hierarchical Clustering

- It delivers as many partitions as there are nodes: so it is not always easy to determine which partition to chose
- Results usually depend on the similarity measure and on the criterion adopted to compute the similarity of the groups, so we may need to do a lot of tries to get the wanted result based on different similarity measures
- It is rather slow, and networks with millions of nodes are out of reach

48 COMMUNITY DETECTION

There are many different methods to identify a community.

49 BRIDGE REMOVAL



This is a **divisive method**, we remove the links that connect the clusters to each other (the bridges) until a cluster is isolated. The problem would then be solved by finding the connected components of the resulting disconnected graph. We can identify bridges via measures that take larger values on those links rather than on internal links.

49.1 Girvan Newman Algorithm for Bridge Removal

The measure for the bridge identification is the link betweenness. Bridges are expected to have large betweenness values because shortest paths between nodes in different communities run through bridges and instead, internal links are expected to have lower betweenness values, because there are many alternative routes going from one node of the community to another, due to the high density of links inside the clusters.

The Girvan-Newman algorithm is a divisive hierarchical clustering method, as it delivers hierarchical partitions by breaking clusters until only singletons remain.

The problem of this algorithm is that it is **quite slow**, it is not practical for large networks. **The bottleneck is the recalculation of the link betweenness.** Faster variants have been developed, for instance computing approximations of the link betweenness scores by using only a sample of randomly selected pairs of nodes, or adopting alternative measures to identify bridges, which are quicker to compute.

The method delivers a full hierarchy of N partitions, **which ones are meaningful and how can they be selected?**

49.2 Quality Function (How to determine how good a partition is?)

We can determine this with quality functions. The idea is that we have a dendrogram, for every partition in the dendrogram we calculate a function, and when this function is maximized we can conclude that the quality of that partition in the dendrogram is optimal inside that dendrogram according that quality function.

We still have issues though:

- The internal link density of the clusters is not enough to assess their community quality. **Random networks have no communities**, so any subnetwork of a random network is not eligible as community no matter how dense the subnetwork is internally, since it won't be connected by a natural similarity, and they have no natural clustering topology.
- It is necessary to distinguish actual communities, where there are high concentrations of links because of specific features of their nodes (similarity), from **pseudo-communities**, where high concentrations of links are produced by chance in the construction process of the network

50 MODULARITY

The idea is to evaluate communities with respect to a **random baseline**, where the baseline is randomized versions of the original network, preserving its degree sequence, just as the **configuration model**.

For each community of a partition, **modularity computes the difference between the number of internal links in the community and the expected value of this number in the set of randomized networks**.

If the network is random (as a result of maybe an Erdos-Reyni model), the modularity of any partition is supposed to be low, because the number of internal links of any cluster of the partition should be close to the expected value in the randomized networks. If the number of links within the clusters is much larger than its expected random value, it is unlikely for such a concentration of internal links to be the result of a random process, and modularity can reach high values.

50.1 Homophily Test and Modularity

The Homophily Test says that if p is the fraction of nodes in group A and q the fraction of nodes in group B, then we have a signal of homophily if the number of actual cross groups edges $< 2pq$.

We can use this same principal for modularity but:

- we can have more than two groups
- we need to make our comparison with a **degree preserving randomization**, like with the one we get from the Configuration Model

50.2 Formula of Modularity

$$Q = \frac{1}{L} \sum_C \left(L_C - \frac{k_C^2}{4L} \right)$$

- L = number of links in the network
- L_C = number of internal links in community C
- k_C = degree of community C
- $k_C^2/4L$ = expected number of internal links in community C

We have a community C , and the number of internal links inside of this community (L_c). We subtract to the number of internal links the number of expected links in the random network, and this number is the degree of the community c to the power of two, divided by $4L$. We sum then the value inside of the brackets for all the communities C , and then we normalize this sum by L , which is the total number of the links in the network.

If $Q = 0$, it means that the number of internal links L_c is equal to the expected number of internal links, and this means that the real community is not a real community but a pseudo-community since we have exactly the same community that we get in the random network. **If $Q = 0$ we have a random network.**

If $Q > 0$ it means that the number of actual internal links is higher than the number of internal expected links in the random network, **which means that the higher is Q , the more relevant communities are to be found, so the idea is to maximize Q for all the partitions that we have.**

Origin of $k^2 C / 4L$

Random links are formed by matching pairs of stubs (half-links) chosen at random. The total number of stubs attached to C is K_c . The probability to select one of those stubs at random is $K_c/2L$ because $2L$ is the total number of stubs of the network (each link yields two stubs). For a random link to connect two nodes in the same cluster C , two stubs must be selected randomly from the cluster C , and this probability is the product of the probabilities of selecting each one: $P_c = (K_c/2L) * (K_c/2L) = k^2C / 4L^2$. Since there are L links in the network, and each has a probability P_c to end up within C , the expected number of internal links in C is $k^2C / 4L$

50.3 Features of Modularity

We want to use modularity to find the best possible partition that we can based on the set of partitions we start with.

- $Q < 1$ for every partition of any network
- $Q = 0$ for the partition in which the whole graph is one community
- Q can be negative and that means that we have partitions in N groups of one node each
- For most networks, Q has a non-trivial maximum between 0 and 1

When Q is close to 0 we can assume that the network is similar to a random graph with no communities, and when is close to 1 we can assume that the clusters are pretty different from what to expect in random networks.

50.4 Modularity Optimization

Modularity was introduced to provide a criterion to choose the best partition out of those found via the Girvan-Newman algorithm, but if modularity is reliable, why not maximize it directly? We could do this with **modularity optimization**: which requires to find the maximum of Q in the space of all possible partitions of the network into communities, and this is a pretty hard problem.

50.5 Newman's Greedy Algorithm

The procedure of this algorithm is:

- To start by partition with one node in each community
- Merge the pair of groups of nodes that yields the highest increase (lowest decrease) of Q
- Continue until all nodes are in the same community
- Pick the partition with largest modularity

This procedure has its **limits**:

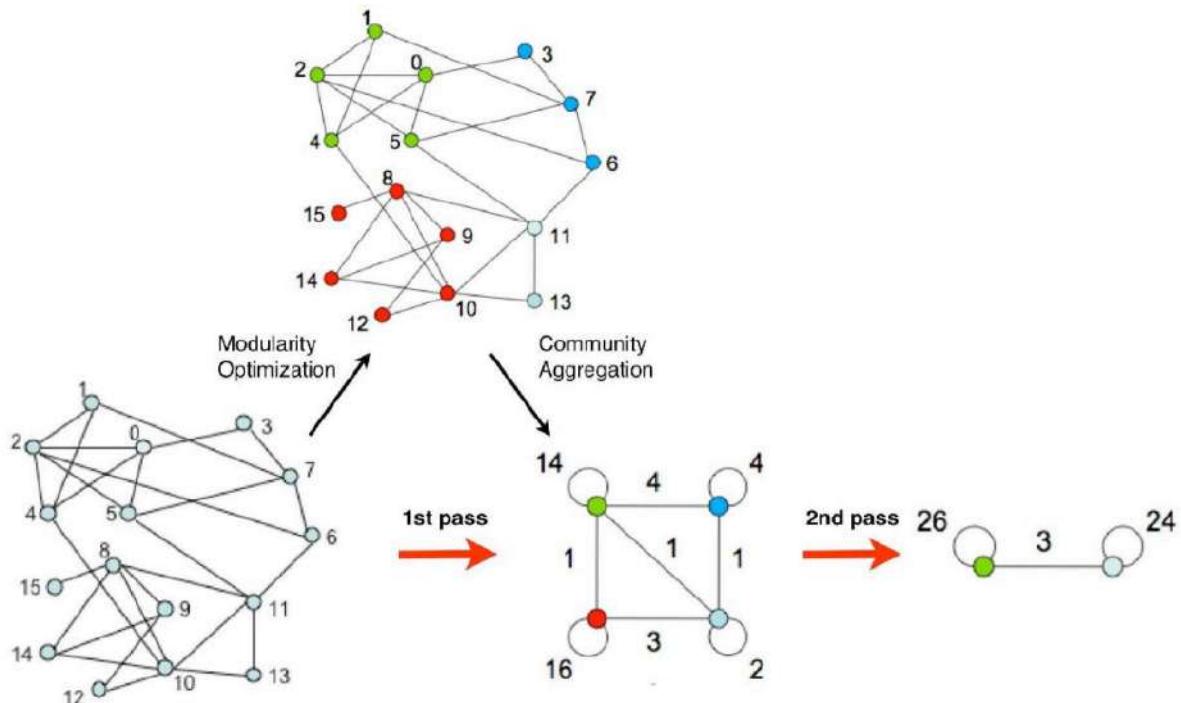
- The method is greedy, since it tries to maximize the modularity at each step. As such, it is likely to get stuck on solutions with sub-optimal modularity
- It tends to generate unbalanced partitions, with some clusters much larger than the others. Because of that, the method is not very fast
- Merging groups of similar size or merging more than two groups at a time mitigates the problem

50.6 Louvian Algorithm

One of the best algorithms for community detection based on modularity optimization. It has a computational complexity that is linear with the number of nodes.

The procedure of this algorithm is:

- 1. We start with each node assigned to a different community
- 2. Loop over the nodes: each node is put in the community of the neighbor that has the largest modularity increase (ΔQ) compared to the current partition. All nodes are revisited over and over until it is no longer possible to increase Q by moving a node to a different community
- 3. The network is transformed into a weighted supernetwork, where each community in the partition from step 2 is replaced by a supernode, links between supernodes are weighted by the number of links joining nodes in the corresponding groups, and links joining nodes in the same community are represented as a self-loop from the corresponding supernode to itself, with weight equal to the number of internal links
- 4. The procedure stops when no further grouping of the clusters in the current partition increases the modularity



50.7 Considerations on the Louvain Algorithm

It has its limits:

- Like Newman's algorithm, **it is a greedy method**, in that it tries to find the best modularity partition at each level of agglomeration. Therefore it yields solutions with sub-optimal modularity
- The final partition depends on the order in which nodes are visited

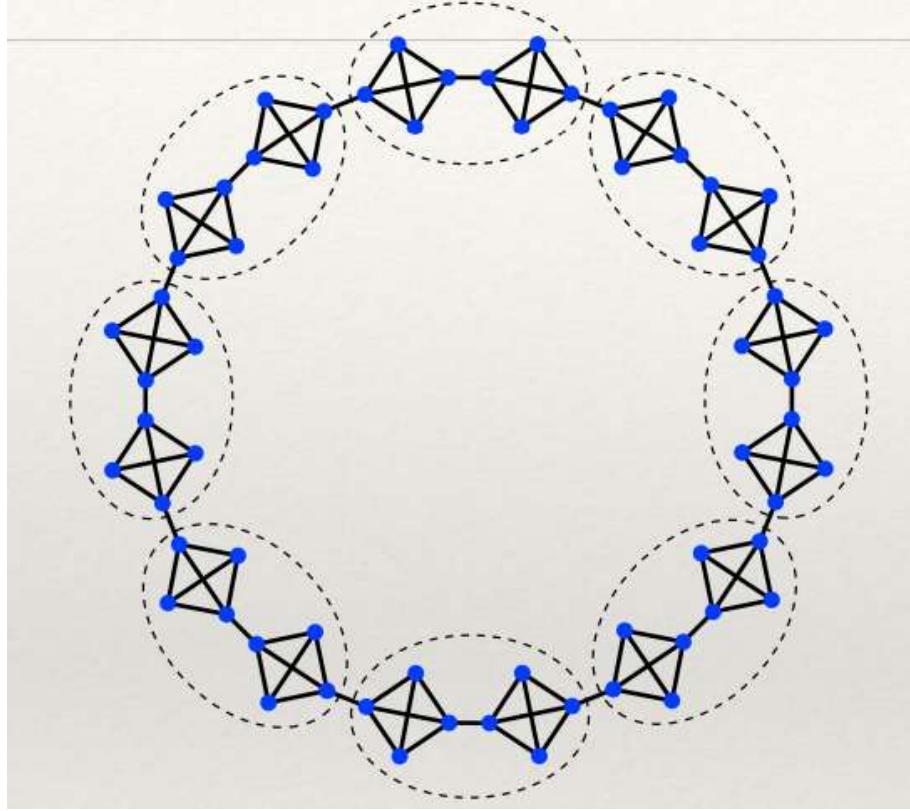
But also a big advantage:

- **The algorithm is very fast because**, after the first iteration, the successive transformations shrink the network very quickly and typically only a handful of partitions are generated. It can detect communities in networks with millions of nodes and links

50.8 Limits of Modularity Optimization

- The maximum modularity tends to be larger on larger networks, so the measure cannot be used to compare the quality of partitions across networks
- The maximum modularity of random networks without group structure (like with the Erdős–Renyi model) can attain large values
- The maximum modularity does not necessarily correspond to the best partition. Communities smaller than a certain size may not be detected, and this is called **resolution limit**.

50.9 Resolution Limit



The natural partition is the one where the communities are the cliques but modularity is larger for the partition where pairs of cliques are merged

A possible solution for the resolution limit is tuning the resolution of the method by inserting a parameter in the modularity formula (multiresolution modularity optimization), but we have to problems with this:

- It is very **computationally intensive** since modularity has to be optimized for multiple choices of the resolution parameter
- A criterion is needed to decide which value of the resolution parameter is most suitable for a given network

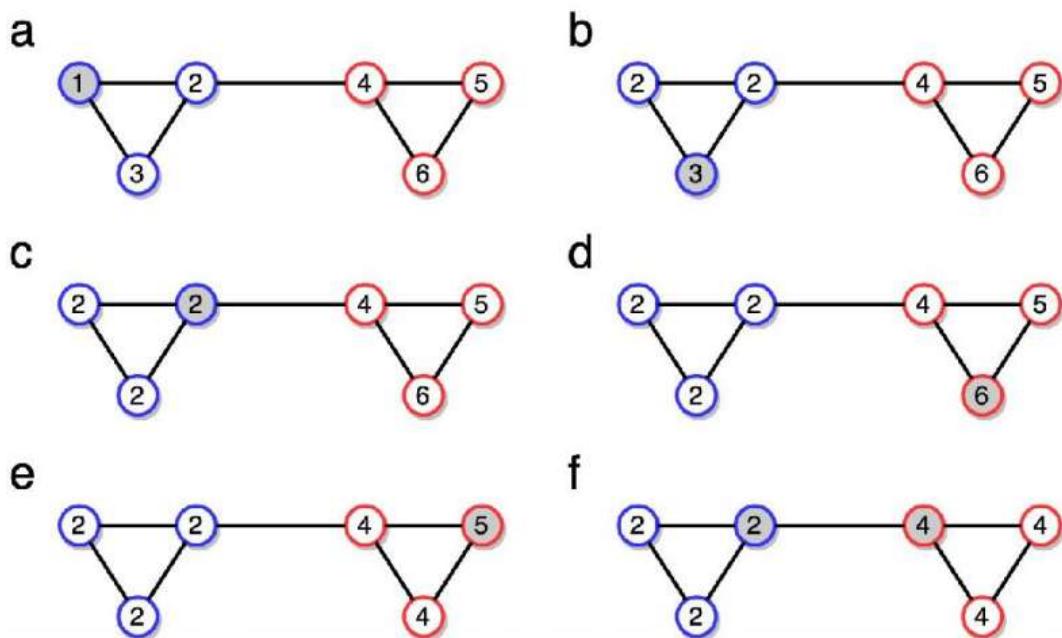
51 LABEL PROPAGATION

The idea is that the neighbors of a node usually belong to the same community. We can suppose that the majority of the nodes of the network belong to the same communities.

The procedure is this:

- 1. We start by having each node assigned to a different community. Each node is given a different label, where every label identifies a different community
- 2. A sweep is performed over all nodes, in random order, so each node takes the label shared by the majority of its neighbors. We simulate a dynamical process where each node is somehow influenced by the majority of its neighbors, we assume that the nodes that we randomly select belongs to the same community where the majority of nodes belong. If there is no unique majority, one of the majority labels is picked at random
- 3. If every node has the majority label of its neighbors, we are then in a **stationary state** and we can stop, otherwise we repeat the second step. The procedure is repeated until we reach a stationary state.

Communities are defined as groups of nodes having identical labels in the stationary state.



Labels propagate during the process, which mean that most labels disappear, others dominate. The algorithm usually converges after a small number of iterations, independent of network size. In the final partition, each node has more neighbors in its own community than in any other, so each cluster is a **strong community**. This algorithm is quite performant since it runs in the order of the number of the nodes present in the network

51.1 Limits of Label Propagation

- The algorithm does not deliver a unique solution, but the outcome depends on the order in which the nodes are visited in each sweep.
- Different partitions are also the result of the many ties met along the process, which can be broken in different ways depending on the sequence of random numbers.
- Despite these instabilities, the partitions found by label propagation in real networks tend to be similar to each other. For more robust results, one can combine the solutions obtained from different runs of the procedure and try to have a better fit for the membership of nodes that fluctuate from one community to another assinging this node to another community with a majority rule.

51.2 Good Things about Label Propagation

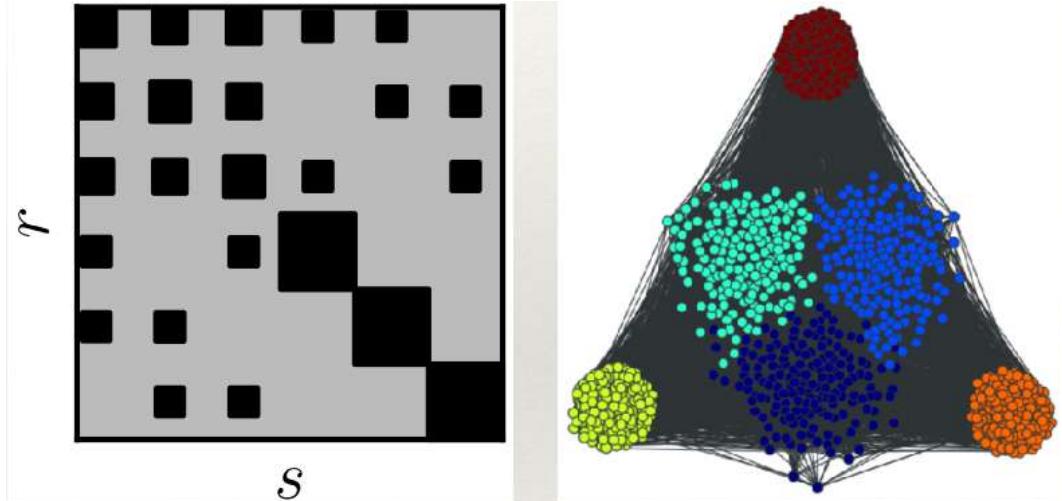
- The algorithm does not need any information about the number and the size of the communities
- It is parameter-free
- **It is simple to implement and very fast**, networks with millions of nodes and links can be partitioned this way
- If community labels are known for some of the nodes, they can be used as seeds in the initial partition

52 STOCHASTIC BLOCK MODELING

The idea behind this model is that we assume that the network has been generated by some model, like the Erdos-Renyi, Barabasi-Albert, Configuration Model etc or any model that reproduces community structures. We know that models are characterized by one or more parameters, so the idea here is not to find communities, but we want to create an artificial network that has communities and then we want to fit our original network with this artificially created networks in order to find the parameters that better explain the formation of communities in the artificial network. More precisely we want to find the for values of the parameters does the model produce networks that most closely represent the one original one that we are investigating.

An example could be starting to create a random network and if we assume that it has been generated by and Erdos-Reyni model, we want to find out for which value of the link probability we get graphs most similar to the original network.

The focus here is community structure, so we need to look for models that generate networks with built-in communities. We start with **Stochastic block models (SBM)** that are the most important class of models for generating networks with communities. The principle of the SMB is that nodes are divided into groups and the probability that two nodes are connected is determined by the groups to which they belong.



Let's suppose that we have **q groups**, and this groups are just one of the possible partitions of our network. G_i is the label of the group of node i . If we want then to estimate the probability that a link between nodes i and j is formed we need to compute $P^*_{g_i g_j}$ where g_i is the group of i and g_j is the group of j . The idea is that we have always the same probability for any nodes belonging to group i and j , but the probability can change when the group changes. The **stochastic block matrix** is the $q * q$ matrix whose element k_l is the link probability P_{kl} . The **diagonal element** P_{kk} is the link probability between pairs of nodes in group k , so its the internal link probability. We can create communities just based on the values of this link probabilities following this method. **With this method we can create randomly a network that is made of different communities.**

- For every row and column that goes from 1 to q , where q is the total number of groups, if the number of row and column is different ($r \neq s$), which means that we are then looking at the cross-group link probability, and we have $P_{rr} > P_{rs}$ (so the internal probability is greater than the external probability), we have community structure, which column links are more likely within than between blocks
- For every row and column that goes from 1 to q , where q is the total number of groups, if the number of row and column is different ($r \neq s$), and we have $P_{rr} < P_{rs}$, we have multipartite structure, which means that links are more likely between than within blocks
- If $q = 2$ and $p_{11} > p_{12} > p_{22}$, we have core-periphery structure, which means that the nodes in the first block (core) are relatively well-connected amongst themselves as well as to a peripheral set of nodes that interact very little among themselves
- If $P_{rs} = p$, for every row and column that goes from 1 to q , where q is the total number of groups, we recover the classic random network, which means that any two nodes have identical probability to be connected; there is no group structure

52.1 Fitting a Stochastic Block Model to a Network

We want to fit the model to the considered original network, and we do this by maximizing the likelihood that, for a given partition of the network, a stochastic block model reproduces the placement of links between the nodes. Of course, the best partition is the one corresponding to the largest value of the likelihood.

A **problem** that arises is that the standard stochastic block model does not describe well real networks, because it ignores degree heterogeneity, which means that networks generated with the model usually have nodes with similar degrees. The solution of this problem is to move to another class of stochastic block models called **degree-corrected stochastic block model (DCSBM)** that uses the actual degrees of the nodes of the network, so we are adapting the configuration model with a stochasting block modeling approach.

To fit the stochastic block model we define the probability that a network G is reproduced by the DCSBM based on a given partition g of the network G 's nodes into q groups is expressed by the **log-likelihood**:

$$\mathcal{L}(G|g) = \sum_{r,s=1}^q L_{rs} \log \left(\frac{L_{rs}}{k_r k_s} \right)$$

- L_{rs} = number of links between group r and group s
- k_r = degree of group r (sum of degrees of nodes in r)

The likelihood of capital G given g is equal to the sum, for every pair of row and column up to q , of the number of links between group r and s , multiplied by the logarithm of the number of links between group r and s divided by the degree of the group r multiplied by the degree of the groups.

The procedure for fitting a stochastic block model to a network is this:

- 1. We start with a random partition in q clusters
- 2. Iterate moving a node from one group to another, selecting at each step the move that will most increase the likelihood (or least decrease it, if no increase is possible), under the constraint that each node may be moved only once, this is still a greedy approach.
- 3. When all nodes have been moved, inspect the partitions through which the system passed from start to end of the procedure in step 2 and select the one with the highest likelihood
- 4. **Stopping criterion:** the likelihood for two consecutive iterations is the same, which means it cannot be increased any further

52.2 Limits of Stochastic Block Modeling

- It is necessary to provide as input the number of clusters, which is usually unknown. In fact, a straight maximization of the likelihood over the whole set of possible partitions yields a trivial division into N groups of one node each. However, there are techniques to estimate the number of clusters
- The greedy algorithm for the maximization of the likelihood gets stuck at suboptimal solutions. To improve the result, it helps to run the algorithm several times with different random initial conditions and select the partition with highest likelihood across all runs

53 METHOD EVALUATION

The problem we want to solve here is how we can tell how good a clustering algorithm is. The solution is that the natural way to evaluate a method is to check whether it is able to find clusters in **benchmark graphs**, that is networks known to have a natural community structure.

We have two classes of benchmarks:

- **Artificial Benchmarks**, which are computer-generated networks, built via some model. From this we will see and understand how our community detection algorithms will detect this artificially created clusters.
- **Real Benchmarks**, which are real networks in which the communities are suggested by the history of the system or by attributes of the nodes, of which we know which are the natural clusters

53.1 Planted Partition Model

Stochastic block models are often used to generate artificial benchmarks, and a special version of these models is used in method evaluations, and this is the planted partition model. The good thing is that the **planted partition model** has only **two link probabilities**: the probability P_{int} (internal probability) nodes in the same community are connected and the probability P_{ext} (external probability) that nodes in different communities are connected.

If $P_{int} > P_{ext}$ we have that the groups are communities, as two nodes are more likely to be connected if they are within the same group than if they are in different ones.

- **q groups of identical size N/q**

- **Expected internal degree of the nodes:** $\langle k^{int} \rangle = p_{int} \left(\frac{N}{q} - 1 \right)$

- **Expected external degree of the nodes:** $\langle k^{ext} \rangle = p_{ext} \frac{N}{q} (q - 1)$

- **Expected total degree of the nodes:**

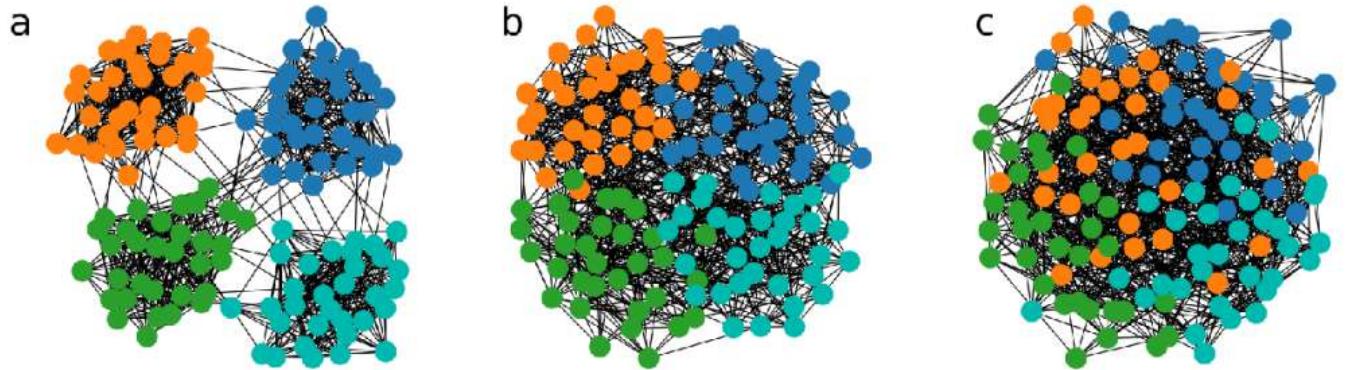
$$\langle k \rangle = \langle k^{int} \rangle + \langle k^{ext} \rangle = p_{int} \left(\frac{N}{q} - 1 \right) + p_{ext} \frac{N}{q} (q - 1)$$

- We can estimate the expected internal degree K_{int} by multiplying the internal probability with $(N/q - 1)$, which is every node in the community itself.
- We can estimate the expected external degree K_{ext} by multiplying the external probability with all the other nodes in the other communities $(q - 1)$ and for all the other external community we have N / q nodes.
- The expected total degree of the nodes is then $K_{int} + K_{ext}$.

53.2 Girvin Newman - GN Benchmark

Is a specific implementation of the planted partition model, in which the network size, degree of the nodes, and number and size of the communities are set to particular values: $N = 128$, $q = 4$, $= 16$, and since $31P_{int} + 96P_{ext} = 16$, hence P_{int} and P_{ext} are not independent parameters.

GN benchmark networks are constructed with a procedure similar to the one adopted for Erdos–Renyi random graphs: we go over all pairs of nodes and connect each with probability P_{int} or P_{ext} , depending on whether or not the nodes are in the same community



- The higher the expected external degree and the lower the expected internal degree, the more difficult it is to detect the communities since the communities have very big internal cohesion but are loosely connected to each other.
- **Expectation** is that the communities should be detectable as long as $P_{int} > P_{ext}$, that is smaller than 12
- **Surprise**, the detectability threshold is quite a bit lower, around 9, due to random fluctuations in the placement of the links and we call this detectability limit)

This benchmark is not realistic, because the degree of the nodes is approximately the same, and we know that realistic networks have heterogeneous degree distributions, and they deal with communities of the same size, which is also something not always true in real networks.

53.3 LFT Benchmark

The problems are that nodes have approximately the same degree and communities have approximately the same size

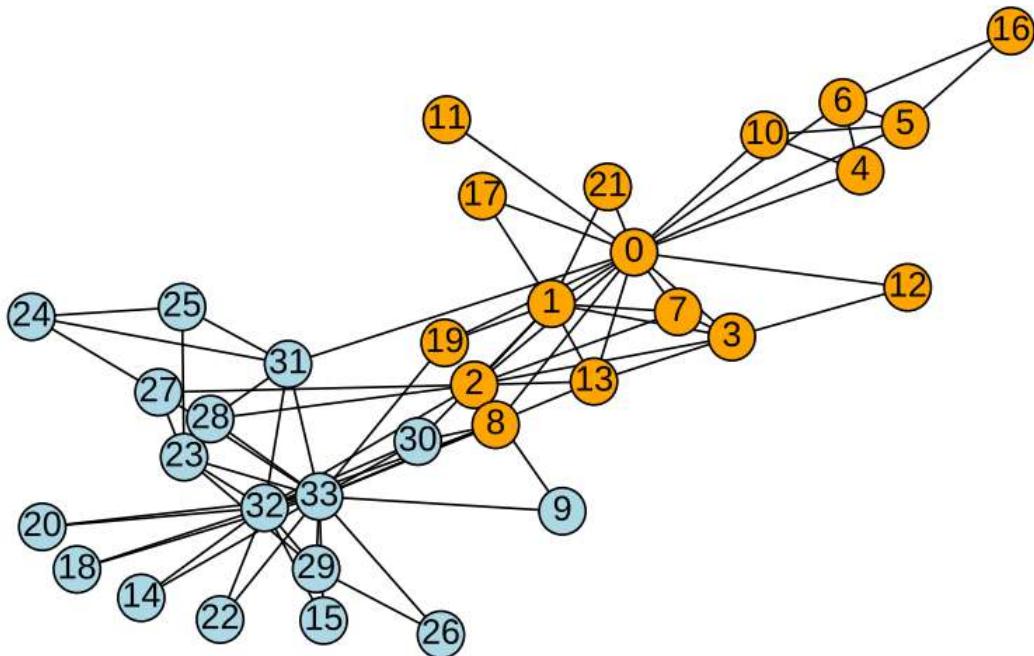
In real networks the distributions of node degree and community size are quite heterogeneous. The **LFR benchmark produces networks having heavy-tailed distributions of degree and community size.**

53.4 Negative Tests

A negative test is to test if the algorithm is finding communities also in networks without communities. If this algorithm finds communities, other than trivial partitions, we can say that the algorithm is not reliable. Of course we can find trivial partitions since one cluster could include everything or we could have N clusters of one node each. Any non-trivial partition would signal the method's inability to distinguish actual communities from subnetworks with high concentrations of links generated by random fluctuations.

53.5 Real Benchmark

A classic example is the Zachary's karate club network, which indicates network of social relationships between members of a karate club in the US. Following a disagreement between the instructor (node 0) and the president (node 33), the club split in two different groups, and we want to identify those groups (identified by the node colors) using a clustering algorithm.



We don't have only the karate club, but a lot of other examples: in many social networks there are groups that users can decide to join, in citation networks papers can be grouped according to their publication venues, Internet routers can be categorized by country, etc.

Do clusters of nodes with similar attributes match the communities found by clustering algorithms based only on the network structure? Well, if nodes with similar attributes are strongly linked to each other, the attributes can be revealed through community detection. If instead the attributes do not play a role in the build-up of the network, they remain invisible to clustering methods

53.6 Partition Similarity

We want to tell how close the partition found by the algorithm is to the planted partition of a benchmark network, and for this we need partition similarity measures.

We have two main classes of measures:

- Fraction of correctly detected nodes
- Normalized Mutual Information

53.7 Fraction of Correctly Detected Nodes

A node is correctly identified if it and at least half of the other nodes in the same community in the detected partition are also in the same community in the benchmark partition. The trick is that if the detected partition has communities obtained by merging two or more groups of the benchmark partition, all the nodes of those clusters are considered incorrectly classified by this measure. The number of correctly detected nodes is then divided by the number of nodes N of the network, producing a number between zero and one.

The problem is that this procedure is arbitrary since we decide if a node is correctly or not classified according to the membership to the cluster that we have in our benchmark and according to a number that oscillates between 0 and 1. This method is easy to implement and is widely used, and it is good enough for easy analysis.

53.8 Normalized Mutual Information

If we know one partition, we can try to estimate how much information we need to infer the other partition. If the two partitions are very similar, little information is needed to transform from one to the other. The more extra information is needed, the less similar the partitions are. **The normalized mutual information (NMI) is a measure that quantifies such additional information.**

We start with two partition X and Y of the same network, and we want to calculate how similar these partitions are.

The probability that a randomly chosen node belongs to cluster x (with size N_x) of partition X can be calculated as the fraction of nodes belonging to X with respect to the total number of nodes in the network, which is $P(x) = N_x / N$

The probability that a randomly chosen node belongs to cluster x of partition X and to cluster y of partition Y is the probability $P(x,y) = N_{xy} / N$, where N_{xy} is the number of nodes shared by x and y

- **Shannon entropy of X:** $H(X) = - \sum_x P(x) \log P(x)$
- **Conditional entropy of X given Y:** $H(X|Y) = \sum_{x,y} P(x,y) \log[P(y)/P(x,y)]$

Based on **Shannon's notion of Entropy**, the entropy of partition X ($H(X)$) is the negative sum of all $P(x)$ multiplied for the log of $P(x)$. While the **Conditional Entropy of X given Y** is the sum over x,y of $P(x,y)$ multiplied by the log of $P(y) / P(x,y)$

The Normalized Mutual Information of X, Y is:

$$\text{NMI}(X, Y) = \frac{2H(X) - 2H(X|Y)}{H(X) + H(Y)}$$

- **NMI = 1** if and only if the partitions are identical
- **NMI has an expected value of 0** if the partitions are independent, as for example when two random partitions are compared
- The problem with this is that detected partitions with more clusters may yield larger values of the NMI even though they are not necessarily closer to the benchmark partition

54 SIGNED NETWORKS

We assume that networks have positive or negative signs that contribute to annotate the edges and the relationships between nodes. They are useful to estimate how balanced is a network. A perfectly balanced network is one that could be cut in two where I have all the friend in one group and all the nodes that belong to different groups are enemies. So a perfectly balanced network is an extremely polarized network, and this can be represented easily with signed networks.

54.1 Segregation and Polarization

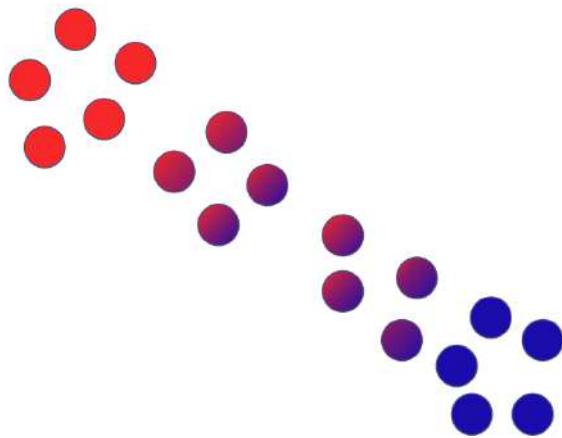
Segregation is based on the idea that the structure of society is shaped in function of immutable characteristics of individuals like ethnic group, age, religious belief or others.

If we make an example based on Juventus or Torino supporters, we can assume if you support one or the other team, your private characteristic is immutable.

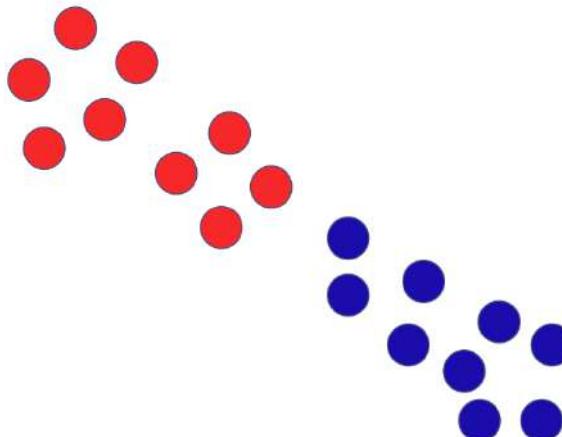
Segregation assumes that you can start with a society that is mixed, composed of juve and torino fans. But once a Torino supporter moves closer to one or more Torino supporter, we can assume that Juve fans will do the same until the two groups are very well distinguishable.

We can observe the segregation phenomena in different ways of structural network methods, we can study the phenomena just looking the node's attribute that refers to some characteristic of an individual, but we can also monitor what happens at the structural level of the network when one of those forces drives the movement or just the link formation between people in the network. We can use as measures high clustering coefficient when people in the same group are tightly connected, this causes homophily, emergence of communities and ghettoization. We are focusing on characteristics that are immutable and since we cannot change this characteristics, we are probably more driven to change the nature of the relationship and links around my self and this in a long term will change the structure of the network itself.

Polarization is different. We can think that polarization is driven from segregation but not only. Polarization is a **state** but also a process, polarization as a state refers to the extent to which opinions on an issue are opposed in relation to some **theoretical maximum**, that is reached when people that show an opinion are tightly connected with each other, and they are opposed to all the other people showing a different opinion. While polarization as a **process** refers to the increase in such opposition over time, so the tendency to reach the theoretical maximum over time.



Our ideas can change over time, and we can be both republican and democrat in a certain moment (if we talk about us politics) but this can change over time and we could become fully democrat or republican.



After a while we can see that we can have a segregation, where people either chose a side or another based on the influence of the people they are connected to.

Polarization and Segregation are similar phenomena, and one can cause the other.

54.2 Issues with Studying Polarization

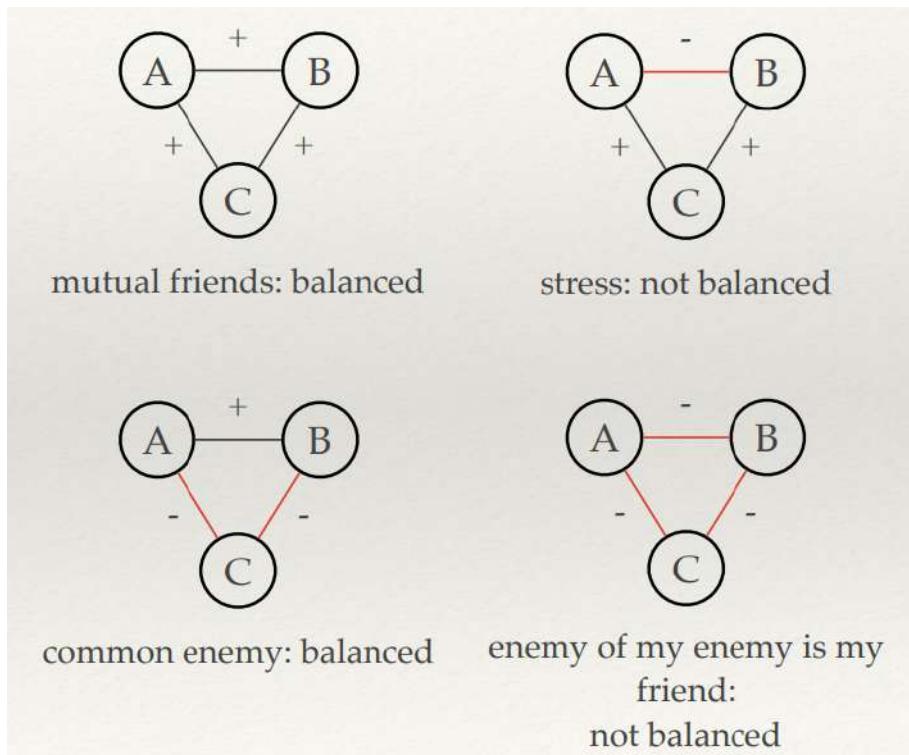
If we have tool to study the structure of a network and to find segregation, we need tools to study polarization as well. When polarization is seen as a **state** we need to understand the opinion of the people, and it can be very difficult in social networks, we could try to analyze by **stance detection** the opinion of someone based on a tweet or post, and while these methods are great they have still quite a bit of flaws, since an opinion of a person is quite difficult to detect. Even the polarization seen as a **process** is difficult to observe since the opinions can mitigate or polarize over time, but people do not necessarily express them, and hence we need to find other ways to understand their opinions and behaviors.

Polarization can even be moved by things such as **selection and influence**, that we studied while talking about homophily, where we can get along with people that share my opinion or we can get influenced by people we get along.

Polarization is difficult to be understood, but an idea is that we need a proper framework to represent friends and enemies, where people go along or people don't, and we can do this with signed networks where we assign positive or negative signs to the edges that connect the persons.

55 STRUCTURAL BALANCE

There are many different domains where we can divide the relationships of an individual as friends and enemies based on conflicts, ideas, trust relationships, taste in something etc.. The **Structural Balance Theory** is a framework to understand the tensions between opposing forces inside a complex system, and this theorem is a good example to understand how local effects can have a domino effect at global consequence at a network level.



This was first formalized in the 40ies. We start with cliques (complete graphs) and we assume that we are aware of the quality of the relationships of this nodes, and edges are identified with a negative or positive sign.

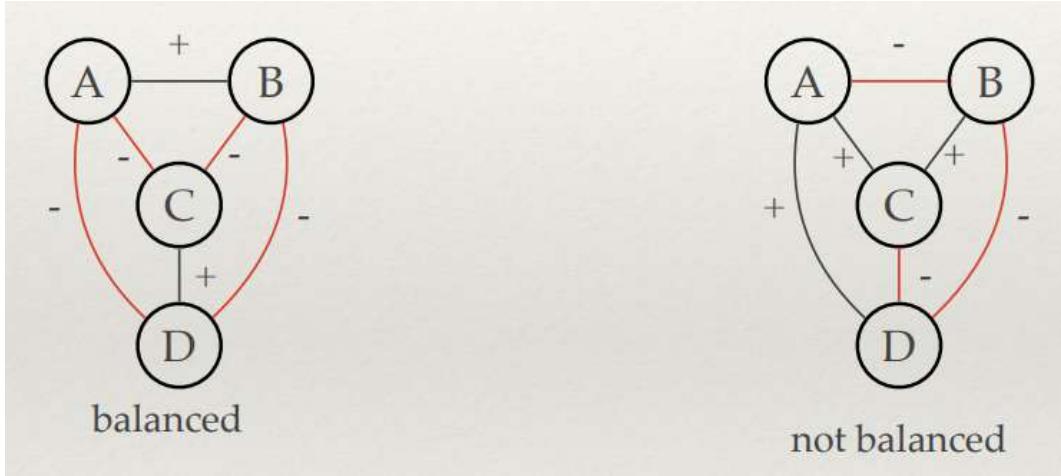
The idea is to understand the cognitive dissonance, which means that we think of many different things that could contradict themselves, and other can agree with other things. This could be a paradox but we are only human and this could happen. This idea of understanding the cognitive dissonance effect was carried on to reasoning on socio-psychological behaviors between nodes.

We have three nodes and we can have certain scenarios:

- When we have a **mutual friend** we have a balanced situation
- When we have a situation where one of our friends doesn't agree with another and we have **stress**, and hence the clique is not balanced
- When me and a friend of mine have an enemy in common we still have a balanced situation
- When everyone in the clique has an **enemy relation** the situation is considered not balanced, since the enemy of my enemy can become my friend

55.1 Structural Balance Properties

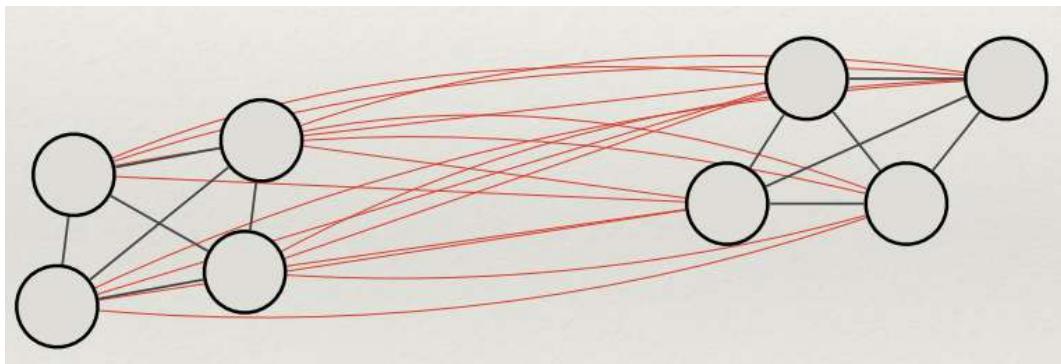
Property 1: A network (clique in this case) is balanced if every triangle in the network is balanced.



Property 1 represents a limit of a social system that has eliminated all the unbalanced triangles, but this is not feasible since we cannot understand all the natures of the links that connect two nodes, but it's a good starting point.

55.2 The Balance Theorem

This theorem gives us a practical way to understand if the network is balanced or not. If we can divide the graph in two groups where nodes are mutual friends inside each group, and enemies are only between nodes in the two groups, then the graph is balanced.



The idea is that we don't want to look for all the triangles in the networks since it is computationally complex, but we just look for two strictly separated groups of nodes that are connected to each other only with negative edges. This has been proved by Harary that proved that this is the only way for a graph to be balanced.

55.3 Proof of The Balance Theorem

Hp.

- a complete labelled graph
- balanced

Ts.

- everyone is friend or I will find two groups X and Y

Proof:

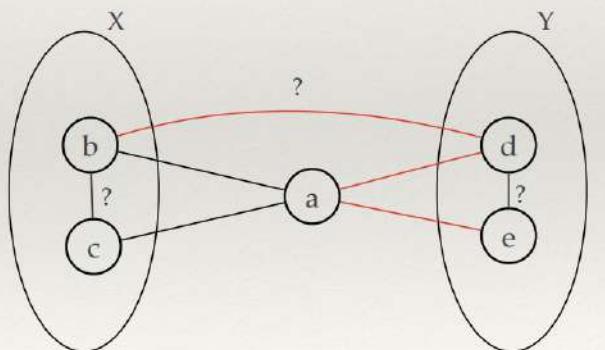
a: node

X: the set of nodes that are **friends** with a

Y the set of nodes that are **enemies** with a

We need to prove that:

- (i) every nodes in X are friends 
- (ii) every nodes in Y are friends 
- (iii) $\forall x \in X, y \in Y: x$ and y are enemies 



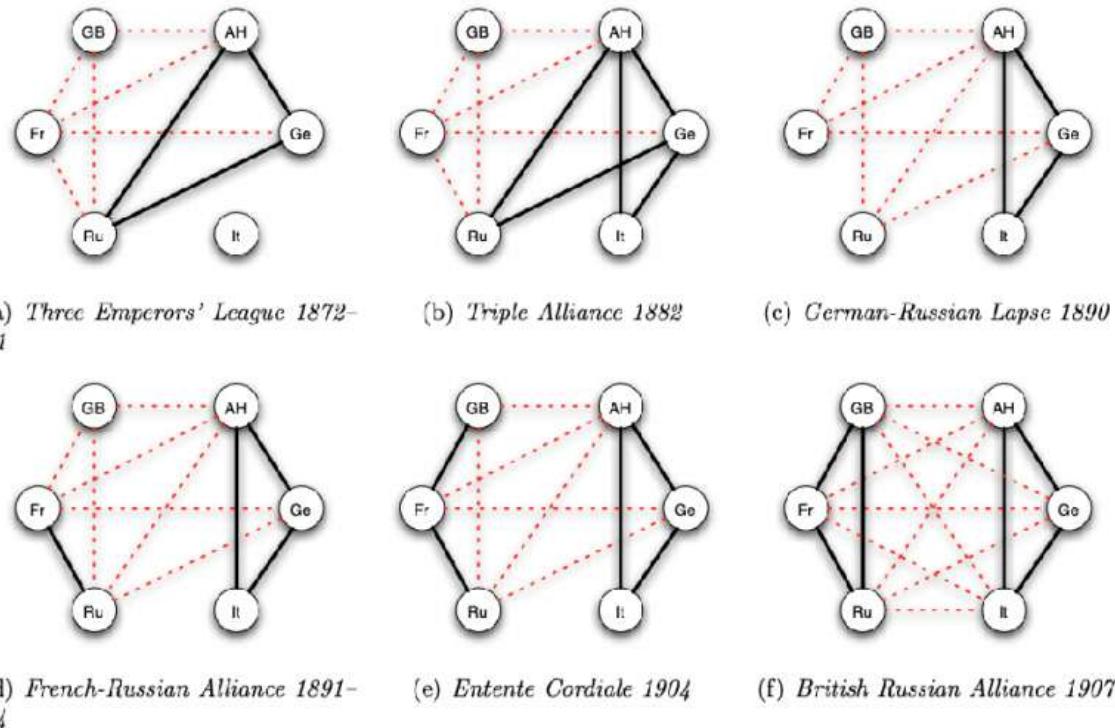
55.4 Applications of Structural Balance

In the real world we don't have boolean difference between balanced or unbalanced network, but we deal with approximately balanced networks, and also generic graphs aren't complete which makes it harder, but we assume that we can apply the balance idea and **we ask how a complete graph might evolve in the search of balance.**

An intuitive idea we use is that physical systems reconfigure themselves continuously to minimize their energy, and we can have an analogy with a complex social system since it wants to balance itself since it wants to minimize the energy of internal disagreement and it moves towards perfect balance.

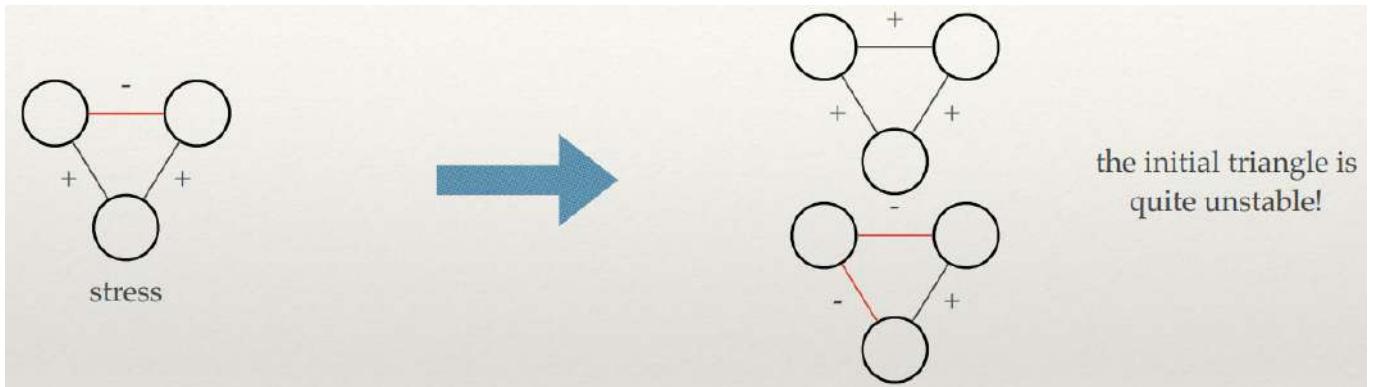
We need to understand though if balance is a good or bad thing, because our environment and psychological thinking makes us think that balance is a positive thing, but in a complex social system we are not always sure that balance is a good feature.

An example is the alliances created between different countries during the 1890s and as new alliances were formed, some states became enemies and some other friends since the alliances could create certain bonds but break others, and as a result a balance is created where 3 countries in a group were enemies with 3 other countries that exploded in world war 1. In international relationships the balances are not a good thing since they can spark wars.

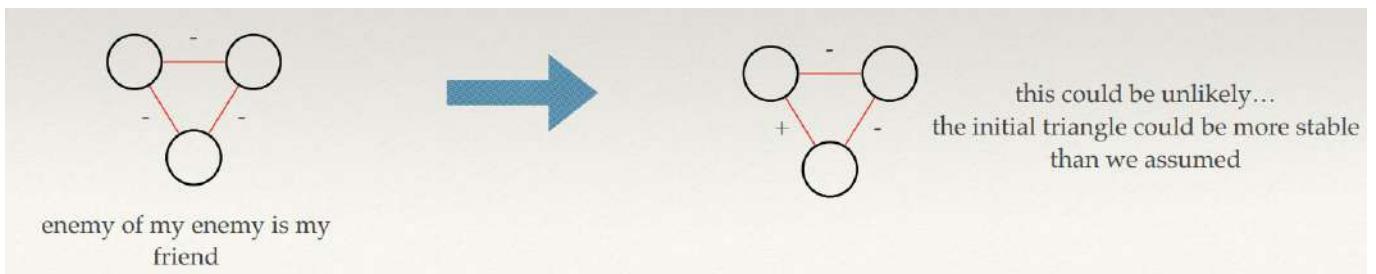


55.5 A Weaker Form of Balance

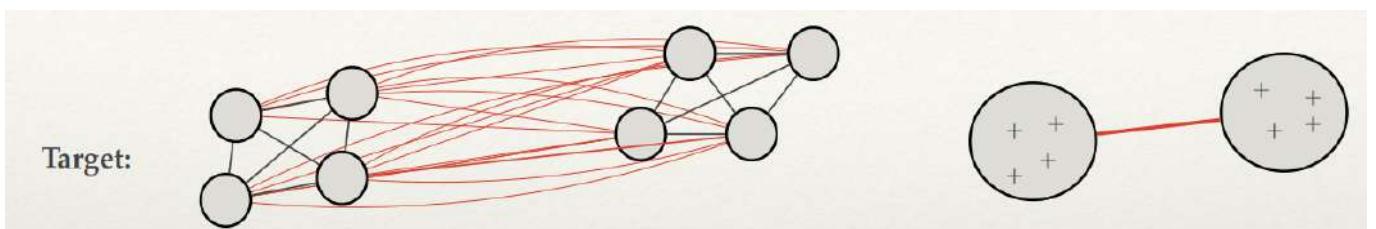
If we think about both the **unbalanced cliques**, in the stress we could either have that the two nodes that were enemies become friend and the stress disappears and we have a balanced situation, or we could have the formation of a disagreement and have a balanced situation. The initial triangle is quite unstable because it could evolve in a different stable situation very easily



If we think about a situation where we **have all negative relationships** we could either have that two nodes become friends and resolve the previous disagreement but this is said to be highly unlikely and the initial triangle could be more stable than we thought. So we can assume that the situation where an enemy of my enemy is my friend is a balanced situation.

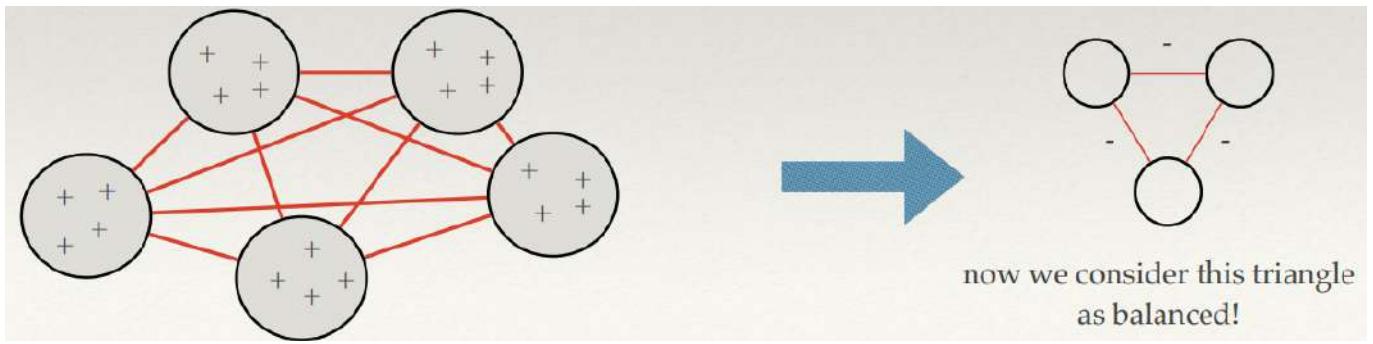


55.6 Weak Structural Balance Property



With the balance theorem, if we find two groups tightly connected with each other positively and negatively to another group, we can deduce that the network is balanced. To evolve this theorem we need to define the Property 2:

There is no set of three nodes such that the edges among them consist of exactly two positive edges and one negative edge. This means that a complete graph, with each edge labeled negatively or positively, is **weakly balanced** if the Property 2 holds.



If we create a network where many groups are enemies with each other, is a balanced network since we consider the enemy of my enemy is my friend triangle as balanced. This means that we can find sets where we have mutual friends inside each group and mutual enemies between the nodes in different groups

55.7 Weakly Balance Theorem and its Proof

The theorem states the characterization of **Weakly Balanced Networks**: *If a labeled complete graph is weakly balanced, then its nodes can be divided into groups in a way where every two nodes belonging to the same group are friends, and every two nodes belonging to different groups are enemies.*

Hp.

- a complete labelled graph
- balanced according P2

Ts.

- If we select any node A , we can build a set X , such that from A 's perspective
 - in X every pair of nodes have a positive relationship
 - in $X + \{A\}$, nodes are enemies with everyone else

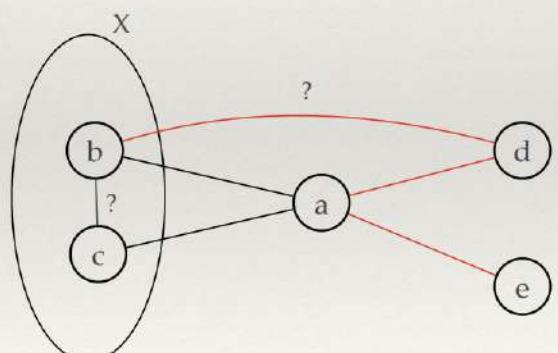
Proof:

a: node

X : the set of nodes that are **friends** with a

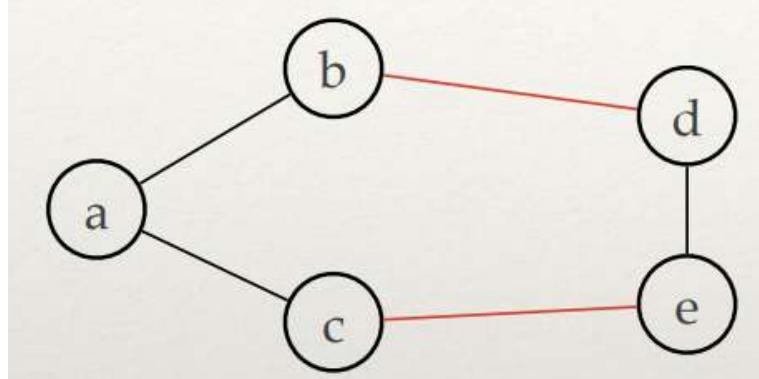
We need to prove that:

- (i) every pair of nodes in X are friends **X**
- (ii) in $X + \{a\}$, nodes are enemies with anyone else **X**



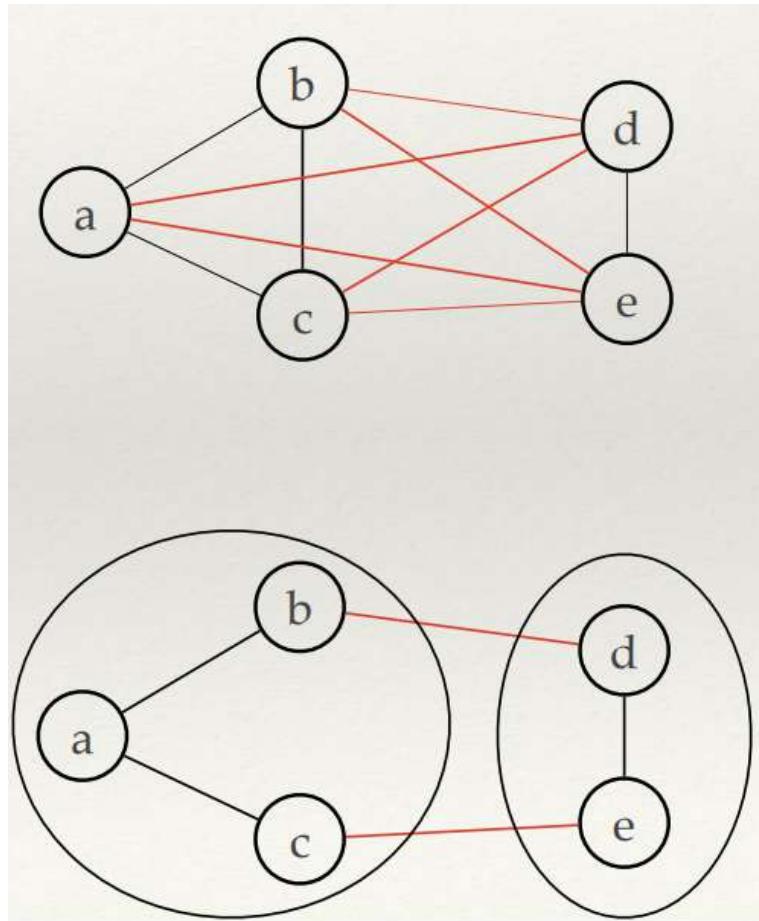
55.8 Generalizing the Definition of Structural Balance

We want to relax the complete graph assumption since it is **highly unrealistic**.



Structural Balance in arbitrary non complete graphs can be achieved with two equivalent approaches:

- Filling the missing edges to achieve balance
- Dividing the graph into two sets



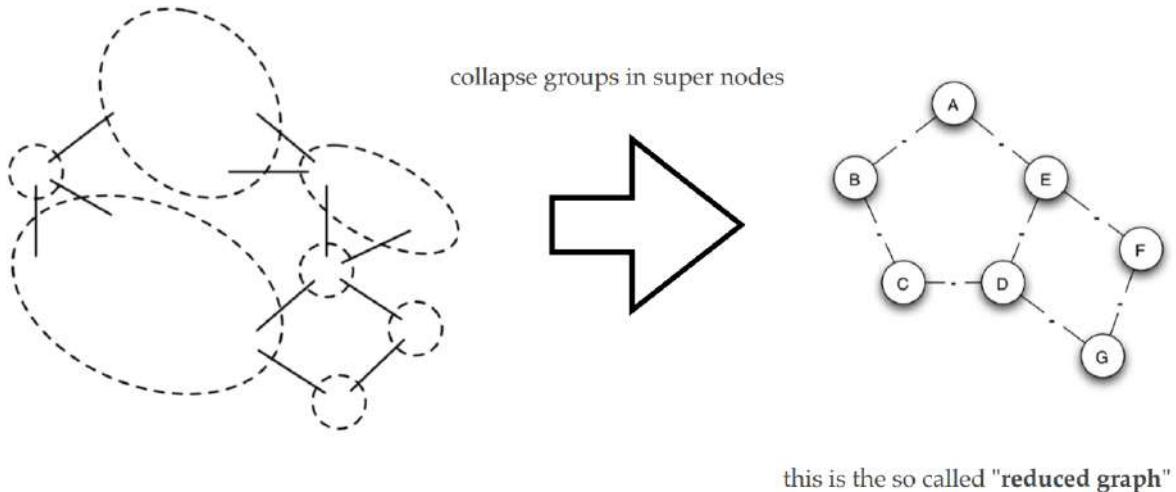
Harary proved a method to check the balance by using a cycle that proves that there is no way of dividing in X or Y. When in a generic graph we find a cycle with an odd number of negative edges we are sure that the graph is unbalanced.

55.9 Claim on Arbitrary Non Complete Graphs

We want to prove a theorem that states that a signed graph is balanced if and only if it contains no cycle with an odd number of negative edges.

To prove this we look for a balanced division between two sets of nodes X and Y, and we do this with a two step procedure:

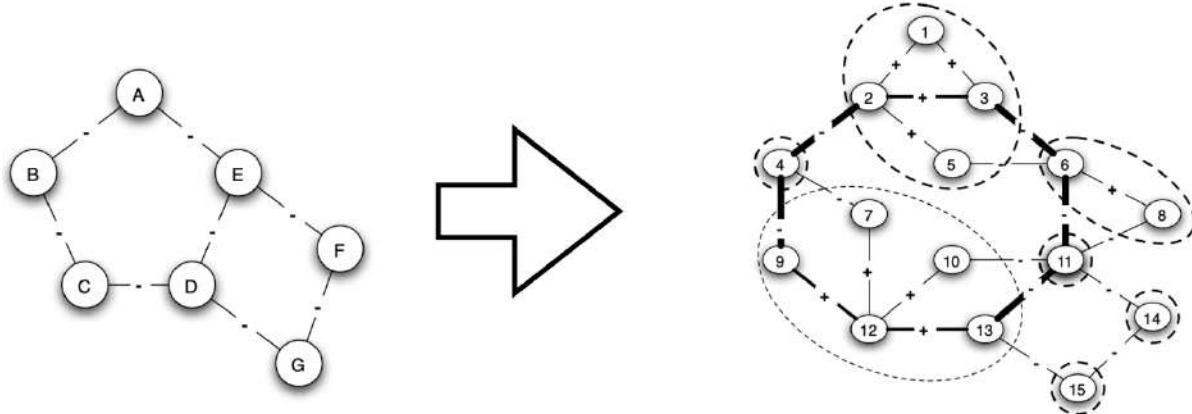
- 1. To convert the graph to a reduced form with only negative edges.
- 2. To solve the problem in the **Reduced Graph**.



Reduced Graphs have only negative edges and we can label the nodes with X and Y, and if this labeling is possible we have a balanced division, otherwise we look for a cycle with an odd number of negative signs. **We want then to prove that we can find such cycles even in the original graph.**

balance: partition between group X and group Y

We have (negative) edges only between nodes in X and nodes in Y



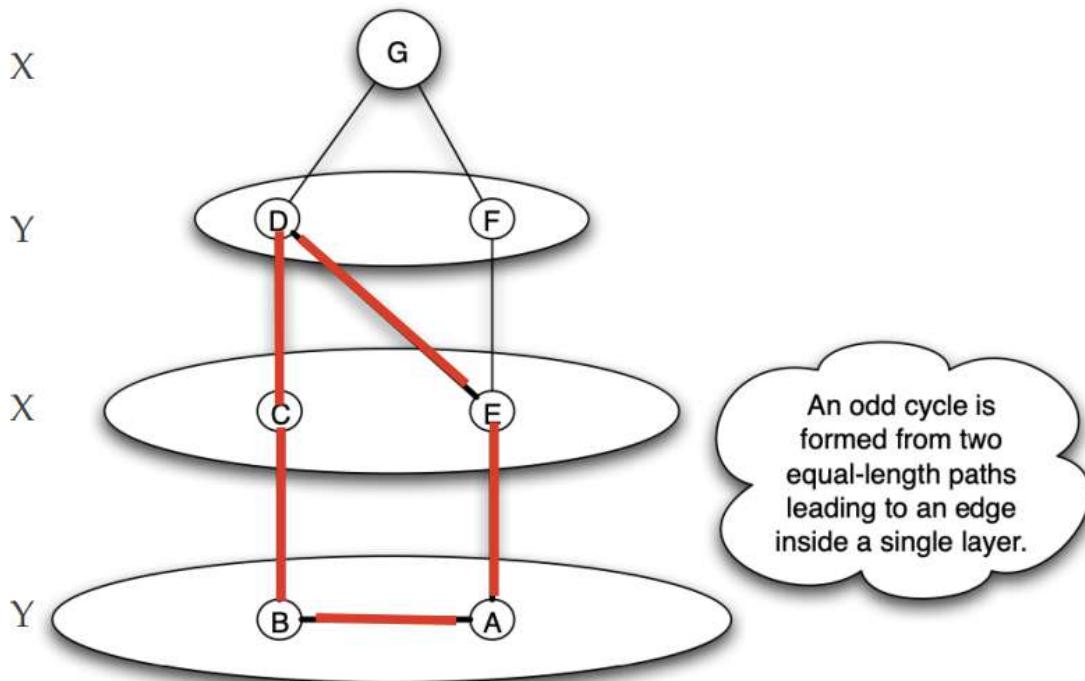
in the reduced graph the **total number of negative signs** in the cycle does not change

Note: a balanced signed graph is equivalent to a **bipartite graph!**

Figure 1: Proof of The Claim

55.10 BFS of the Reduced Graph

All the links in the reduced graph are negative, we start and create the layers etc.. After this we can assign an X to the root, Y to the layer immediately after the root, the other is labeled again with X and the next with Y and if the graph is balanced we can have a perfect partition and this means that we cannot have a negative connection inside two nodes in the same layer, but we are assuming that the graph is unbalanced for the proof.



$$h + h + 1 = 2h + 1$$

that is an odd number!

We can prove that there is a negative cycle in this BFS and that this cycle has an odd number of negative edges. We count the negative edges from one layer to the other and we get the height of the layers h , and if there is two negative links we sum them and them we get $2h$. To this $2h$ we sum the edge that must be connecting two nodes in the same layer if the graph is unbalanced and we just add 1 to $2h$ and we get $2h + 1$ which is an **odd number**.

55.11 Recap of the Proof (Negative Cycle in BFS)

If all edges in the reduced graph connect nodes in adjacent layers of the BFS, then we can label nodes as X and Y and they provide a balanced division of the nodes in the original graph. Otherwise if we cant do that, it means that we have an edge connecting two nodes in the same layer and we produced an odd cycle in the reduced graph, we can then produce a cycle with an odd number of negative signs also in the original graph.

These are the only two possibilities so the graph is either balanced and we cannot produce a cycle with an odd number of negative signs or the graph is balanced and we can produce such a cycle.

56 GAME THEORY

Game Theory gives us a simplified framework to understand how individual strategies can create an intrinsic interdependence in the behaviors of participants to a complex system

56.1 What is a Game?

We are talking about a particular definition of a game, this could be useful in real life in order to set up our strategies when we are exposed to different options and when we have to interact with other agents that each have their own strategies. We are able to define the **connectedness** in complex systems, which is characterized by the **intrinsic interdependencies** between the attributes and behaviours of individuals in a network.

In this situation people don't only connect with each other, but they want to interact with each other and reach a **satisfaction**, given the systems inputs and characterization, but the single individual may want to get a certain satisfaction and sometimes this satisfaction can be influenced by the actions of other individuals and does not only depend on individual's choices.

56.2 Students Example

We have two students that want to pass an exam, and the way of passing the exam is dependent on the actions and behaviours of the other students.

We have assumptions such as:

- they cannot study and prepare the presentation at the same time
- they cannot communicate with each other

The exam is passed if:

- if the student studies for the oral exam he gets a good grade
- otherwise if he doesn't study for the oral he will get a lower grade

The presentation plays this role:

- if one or the other prepare it they both get a good grade
- if neither of them prepare it they will get a lower grade
- if both of them prepare it they will get the highest grade

The final vote is given by the average of the exam and the presentation

We want to understand which is the best option to pass this exam, so we want to formalize this setting to understand if it is better to study or to make the presentation. To understand which is the better solution we use a **payoff matrix**, the idea is that player 1 and 2 want to understand their payoff according to the strategies applied to preparing the exam or the presentation.

If both the players just make the presentation, they won't study and get a lower grade (80) from the oral exam, but they will get the maximum grade from the presentation (100) and hence they will get a grade of 90 each. And so on and so forth for all the possible combinations.

- Two students want to pass an exam

- assumptions:**

- they cannot study AND prepare the presentation;
- they cannot communicate with each other

- Exam:**

- if one study -> s/he gets 92 points;
- if one does not study -> 80;

- Presentation:**

- if one or (xor) the other prepare it: 92 for both;
- if neither of them prepare it: 84;
- if both of them prepare it: 100;

- Final vote:** average on Exam and Presentation

Payoff matrix: it describes the set up

Player 2

		P	E
		P	90, 90 86, 92
Player 1	P	92, 86	88, 88
	E		

P_1 plays P \wedge P_2 plays P \Rightarrow

$$P_1 \text{ gets } 100 \text{ for P and } 80 \text{ for E} \Rightarrow P_1 \text{ gets } \frac{80 + 100}{2} = 90 \\ P_2 \text{ gets } = 90, \text{ too.}$$

56.3 Basic Ingredients of a Game

The games are composed of:

- Players**
- Strategies:** these are a set of options and choices for every player
- Payoff:** this is the outcome for each strategy

These three components put together allow us to compute the payoff matrix, which can be calculated from scratch from every player, hence is a type of knowledge available to all the players. And with these matrix we want to reason about how to players will behave in a certain situation, called game.

56.4 Reasoning about Behavior in a Game

The first thing we need is a **tractable problem** for which we can make some assumptions:

- everything that a player cares of is in the payoff matrix
- everything about the structure of the game is known
- players are rational

The strongest assumption is that the players are rational, since we know that in real life games and problems we are not rational at all, but sometimes we can formalize these ingredients (even the irrational ones) and make them rational and hence have our tractable problem.

If we come back to the exam problem, the player1 can play his decision according to the decision of the second player. A player always needs to compare the payoffs he has in his matrix according to his own decisions and also the other player's decision. Making the best decision for each player results in having a **Strict Dominant Strategy**, and for the example of the students, if they both play their strict dominant strategy, they will both give the exam and wont get the highest possible grade that they would get if they considered another possible strategy. If one player understands that his choice could even benefit the second player he could play this, but could be risky, since the other player can take the advantage but it could be possible that the first player could not get the best outcome.

The solution is that in these situations all the players play with a defensive approach, since they cannot communicate with one another.

56.5 Prisoner's Dilemma

Police have two suspects in custody but don't have any proof, and they ask the suspects to confess. The police says that if they confess they won't make any jail time (0) but the other person will get (9) years in prison. If both of the suspects won't confess, they will both get (1) year in prison. If both of them confess they will get (4) years in prison. The ideal situation for a suspect is to confess while the other is not going to confess.

		Suspect 2	
		Not Confess	Confess
Suspect 1	Not Confess	-1, -1	-10, 0
	Confess	0, -10	-4, -4

Confessing is the best choice for the suspect1 if the other suspect won't confess. Confessing is a strict dominant strategy for both players, but if they both confess they will get the worst outcome, since they both go to prison, and this is counter intuitive.

Of course the opportunity to make 0 years of jail time is not true, but is an incentive the police officers are using to stress the suspects into a confession.

The payoff matrix can be designed in different ways and provide different kinds of outcomes, and it depends of course on assumptions and how certain factors behave (like a good professor or a bad cop)

56.6 Formalization of the Best Responses

- ❖ **Players:** 1, 2 (it can be generalized for more players)
- ❖ **Strategies:** S, T (we can have more strategies)
- ❖ $P_1(S, T)$: Payoff of playing S for P_1 , and T for P_2
- ❖ **Best Response** for P_1 : $\forall S' : P_1(S, T) \geq P_1(S', T)$
- ❖ **Strict Best Response** for P_1 : $\forall S' : P_1(S, T) > P_1(S', T)$
- ❖ For P_2 we have symmetrical definitions

The **best response** for player1 is that for each strategy S where the player1 plays S and player2 plays T is greater or equal to playing every other strategy. The strict best response is when we don't have greater or equal but just greater.

56.7 Formalization of Dominant Strategies

- ❖ **Dominant Strategy:** a P_1 's strategy that is best response to every strategy of P_2
$$\forall S', T : P_1(S, T) \geq P_1(S', T)$$
- ❖ **Strict Dominant Strategy:** a P_1 's strategy that is strict best response to every strategy of P_2
$$\forall S', T : P_1(S, T) > P_1(S', T)$$

- **Dominant Strategy:** for every strategy that can be played by player 1 and 2, we have that the payoff of playing S and T is greater or equal of playing any other strategy.
- **Strict Dominant Strategy:** is the same as the dominant strategy but with only the greater.

56.8 What if only one Player has a Strictly Dominant Strategy?

If we have two companies that want to make a console and one company has a much bigger market reaching than the other, one company will reach a much higher market sale size of the market segments.

In this situation it could be that there is only a strict dominant strategy for the first company and not for the other. This happens because:

- Players must make their decisions simultaneously
- The first company can decide its strategy with no regards of the second companies' move
- Secrecy, since the second company must move without knowing the first company's move, but payoff matrix shares full knowledge of the outcomes
- The second company is subordinate to the first company, and its best strategy is to stay away from the first companies' market segment

56.9 What if no one has a (Strict) Dominant Strategy?

If the payoff matrix doesn't give a dominant strategy, we need another way to predict what is likely to happen during the game.

We can make an example using two companies that have three big clients and use a strategy for each client. **But we can have some events or requests from the clients that might complicate the situation and that give no strict dominant strategy for the companies.**

But inside of the payoff matrix we might have a strategy that has a great outcome for both players, since they maybe don't want to change their strategy because they have no incentive to do so. The system got an equilibrium state, with no force pushing towards a different configuration since it won't give them an higher satisfaction, and this is called the **Nash Equilibrium**.

56.10 Nash Equilibrium

(S,T) is a Nash Equilibrium if S is a best response to T and T is a best response to S.

The notion of equilibrium cannot be derived from rationality since we don't always have a dominant strategy. We are in a situation where no force is pushing towards a different outcome and hence no player have an incentive to change its own strategy.

56.11 Multiple Equilibrium

If we have more than one Nash Equilibrium, the players need to coordinate without any form of communication, and this is a big problem. The solution to this problem was proposed by Shelling and it focuses on **focal points**. The idea is to reason on the natural motivations to focus on one of the Nash Equilibrium, and maybe we can have some **social conventions** that aren't shown inside of the payoff matrix but that can be useful to solve this problem. Then we try to implement inside of the payoff matrix the reasoning that we can make on the **intrinsic features** that can point us to selecting an equilibrium.

56.12 What if Two Players don't Agree

Players tend to disagree with each other, and it is very hard to predict the equilibrium that will be played and hence we need some a priori agreement between the two parts.

56.13 The Stag-Hunt Game

We have two hunters that if they hunt together they can catch a stag, and if they hunt separated they can catch a rabbit each. If they cooperate they get the best outcome which is a stag, if they don't they catch a rabbit. If one cooperate, and the other doesn't, the selfish hunter will get a rabbit, and the other will get nothing. This is similar to the prisoner's dilemma since if they coordinate, they will get the highest payoff, but trying to coordinate is risky. **It could be that the satisfaction of one of the hunters could be the highest if the other doesn't get anything, and even if coordinating is the better choice, it won't always be considered.**

56.14 No Nash Equilibrium

		Player 2	
		Head	Tail
		Head	-1, 1
Player 1	Head	1, -1	-1, 1
	Tail	1, -1	-1, 1

No Nash Equilibria!

If we consider the matching penny game where each player has a penny that can show head or tail, if the two coins match player1 loses otherwise player1 wins. If we compute the payoff matrix we see that we have no Nash Equilibrium.

We can solve these problems with mixed strategies, introducing randomization and probabilities.

56.15 Mixed Strategies

In mixed strategies we model the game using randomization, and the strategies are probabilities that run from 0 to 1. Player1 can choose the strategy S with probability p and hence T with probability $1-p$. Player2 can choose the strategy S with probability q and hence T with probability $1-q$. This means that each player chooses a mixing between the given strategies present in the payoff matrix.

If $p = 0$ it means that Player1 is playing T, while if $p = 1$ it means that Player1 is playing S and these are called **pure strategies**. This is a simple way to have a general framework that includes all the games that do not depend on randomization and probability to have a solution.

56.16 Payoffs for Mixed Strategies

We know that the payoffs are random and we want to understand how to compare them. We can calculate the expected payoffs of the pure strategies of head and tail if we consider the penny example.

From the point of view of the Player1 the payoff of the pure strategy for the Head is $1 - 2q$, while for the Tail is $2q - 1$.

56.17 Equilibrium with Mixed Strategies

If $1 - 2q$ and $2q - 1$ are different we have a **contradiction**, but this is impossible because the two factors are the expected payoff of the pure strategies for Head and Tail, and if these two payoffs are not equal, one would be greater than the other and hence we would have a Nash Equilibrium and we would contradict the assumption of not having it

If they are equal we have the **indifference principle**, and we have that $p = q = 1/2$ and we have that the probability for the matching penny games is $1/2$, which means that half of the time we have heads and the other half tails.

56.18 Interpretation of The Indifference Principle

The **Indifference Principle** states that the choice of p and q are un-exploitable for the other player to decide their strategies

If one player believes that the other will choose head more than half of the times, then he will win more than half of the times by choosing tail. If a player plays randomly head or tail, the randomness is un-exploitable for the other player.

56.19 Optimalities – Pareto Optimality – Social Optimality

We have studied Nash Equilibrium, which creates a situation where each player's strategy is a best response to the other player's strategy, but this does not mean that the players will necessarily reach an outcome that is the best possible. It is possible to classify outcomes not just by their strategic or equilibrium properties, but also by whether they are good for ourselves and the others.

If we want to focus on greater good of society and not only for the best outcome for individuals, we need to change the type of equilibrium that we want to reach.

For example we can look for the **Pareto Optimality**. A choice of strategies, one by each players, is Pareto Optimal if there is no other choice of strategies in which all the players receive payoffs at least as high, and at least one player receives a strictly higher payoff. This means that a mutual agreement is needed to play the pairs of strategies that give out the best outcome.

We also have Social Optimality, which has a stronger definition compared to the Pareto Optimality. A choice of strategies, one by each players, is a **social welfare maximizer** (or social optimum) if it maximizes the sum of the players' payoffs. Social Optimum is also a Pareto Optimum.

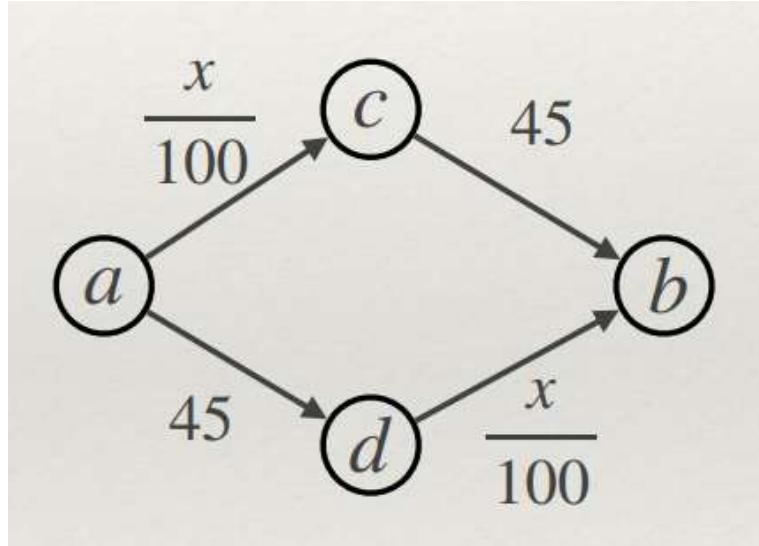
56.20 Networks and Game Theory

We can model traffic models using networks since we have nodes connected with many other nodes, if one agent has to select one connection out of his choices, he needs a strategy to be able to do so, and game theory can help to select the best possible strategy. It is likely that agents will select the strategy that leads to the highest payoff, that means that if we have many agents in this game we can have a **multi-agent system** where each agent will evaluate payoffs according their and everyone's else strategies

Traffic networking is composed by individuals need to evaluate routes in the presence of the congestion that is the results of the decisions that have been made by the agent himself and all the other agents. The models for network traffic may give unexpected results, like the example of the removal of a bridge in korea that when replaced with a river, reduced traffic congesture in the city.

57 TRAFFIC AT EQUILIBRIUM

In this formalization we use directed graphs where **edges** are highways and **nodes** are exits where we can get on or off an highway. As an assumption we consider that every agent wants to go from point a to point b. The **weight** is the travel time in minutes and x is the number of cars (4000 for example)



57.1 Traffic Game

The players are the drivers, hence we have more than two players. Each player's possible strategies are the routes from a to b. The payoff is the negative of a player's travel time, where the faster time is obviously the better strategy, so we want to minimize the travel time.

In games with more than two players:

- The payoff of each player depends on the strategies chosen by all
- The Nash equilibrium is a list of strategies (one for each player), so that each one is a best response to all the others
- With this formalization we can redefine the ideas behind Dominant strategies, mixed strategies, Nash equilibrium with mixed strategies in a way where they all have direct parallels

57.2 Equilibrium Traffic

No dominant strategy in a traffic game, which means that either route has the potential to be the best choice for the player if all the other players are using the other route.

We have **Nash equilibrium**, any list of strategies in which the drivers balance themselves evenly between the two routes. With an even balance, no driver has an incentive to switch over to the other route.

57.3 Can we have a Nash Equilibrium without an even balance?

We prove that we cannot have a nash equilibrium without an even balance. If we have x drivers driving on the upper route and $1-x$ drivers using the lower route. If x is different from $1-x$ then one route is slower than the other, and any driver in the slower route will have an incentive to switch to the other route, and this brings a contradiction because if there is one individual with an incentive to switch we wont have a nash equilibrium and this means that x needs to be equal to $1-x$

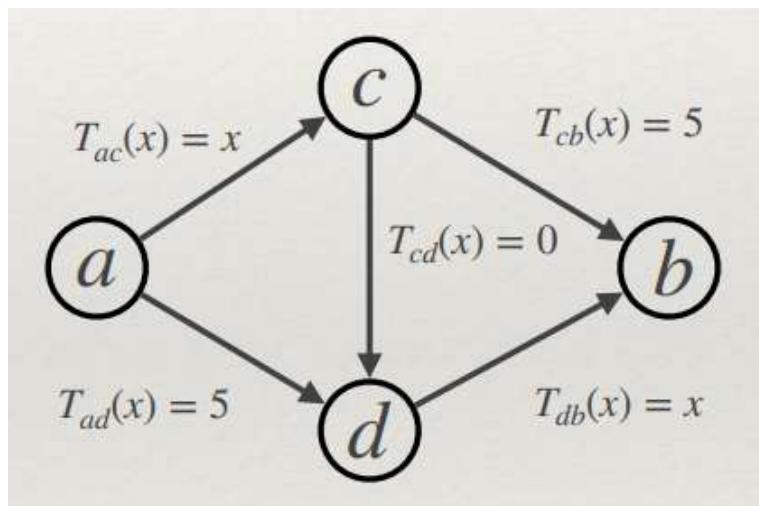
58 BRAESS'S PARADOX

To introduce and formalize this paradox we need to start by saying that **small changes in the network can lead to counter intuitive results**. If we add a certain highway with a very low travel time we can have a new nash equilibrium in our network, since every driver uses the new fast route but then we would have longer travel times due to congestion of the new fast route.

If the new route takes a longer travel time, switching from this route would take still an high amount of time, so the new highway acts like a magnet that attracts all the drivers into this route and hence it degrades its purpose of shortening the travel time. And of course there is no way that a driver gets back to the even balance solution that was better for all the agents in the network, since every driver look at his own self interest in taking the fastest route to get to his destination.

This is not really a paradox, instead it is very similar to the prisoner's dilemma. An intuition could lead us to thinking that upgrading is always a good thing, but in reality it is not true.

58.1 The Social Cost of Traffic at Equilibrium



We want to quantify how far traffic at equilibrium can be from optimal traffic. Each edge e has a **travel time function** $T_e(x)$, which is the travel time through an highway with x cars passing through it.

58.2 Traffic Pattern

A **traffic pattern** (z) is a choice of a path by each driver. The social cost of a given traffic pattern is just the sum of the travel times driven by all drivers when they use this traffic pattern. When a traffic pattern achieves the minimum possible cost we have a **social optimality**, and as we know socially optimal traffic patterns are social welfare maximizers in the traffic game.

We have a **unique nash equilibrium in this game**, where all the drivers chose to drive through this route and hence the equilibrium has a larger social cost. We want now to understand if there always is an equilibrium traffic pattern and if there always exists an equilibrium traffic pattern whose social cost is not much more than the social optimum.

58.3 Finding a Traffic Pattern at Equilibrium

To prove that an equilibrium exists, we need to follow this procedure:

- 1. start from any traffic pattern
- 2. if it is an equilibrium, stop
- 3. else, there is at least one driver whose best response is some alternate path providing a strictly lower travel time
- 4. pick one of these drivers, and go to step 2.

This algorithm is based on **best-response** dynamics where each driver changes its strategy according to their subjective best response. We need of course to show that best response dynamics will stop.

58.4 Does a best-response dynamics always stop?

No, in the Matching Penny game it will run forever because it lacks of an equilibrium, with pure strategies of course. In principle, even in the traffic game we can have a best-response dynamics that could run forever if we do not have an equilibrium.

We can prove that in the traffic game the procedure stops, proving as well that an equilibrium exists and that an equilibrium can be reached by a process in which drivers constantly update what they are doing according to best response.

58.5 Progress Measure

To check if the best-response dynamics will eventually stop, we need a **progress measure** to track the process and to determine how far we are from the process to stop. If a driver wants to understand if changing its strategy will lead to a better outcome compared to another, we need a progress measure. And we need this progress measure also to better understand if our social cost is increasing or decreasing, since if we consider a selfish behaviour, we don't pay attention to the social cost.

The social cost of the current traffic pattern is not a good progress measure though, since some best-response updates by drivers can make the social cost better, but others can make it worse. The social cost of the current traffic pattern can oscillate, and the relationship with our progress toward an equilibrium is not clearly defined.

We need hence forth to define an alternate quantity that strictly decreases with each best-response update, and for this we use the potential energy.

58.6 Potential Energy

The potential energy of an edge e is the total travel time experienced in this edge when we have 1,2,..,x cars passing through this edge. If an edge e has no drivers on it it has a potential energy of 0.

The potential energy of a traffic pattern w is the sum of all the the potential energies of all the edges, with the current number of drivers in this traffic pattern.

$$\text{Energy}(z) = \sum_{e_i \in z} \text{Energy}(e)$$

The energy of e is different from $xT_e(x)$ but it is a sort of cumulative quantity instead, in which we imagine drivers crossing the edge one by one, and each driver only experiences the delay caused by themselves and the drivers crossing the edge in front of them

58.7 Does the best-response dynamics stop?

If we prove that best-response dynamics will stop, then we have proved that an equilibrium always exist. If we prove that the potential energy strictly decreases at each step, then we have proved that best-response dynamics stops. When a driver abandons one path in favor of another, the change in potential energy is exactly the improvement in the driver's travel time.

Let's recall that the potential energy of edge e with x drivers is:

$$\text{Energy}(e) = T_e(1) + T_e(2) + \dots + T_e(x-1) + T_e(x)$$

When one of these drivers leaves it drops to:

$$= T_e(1) + T_e(2) + \dots + T_e(x-1)$$

Summing up, $\text{Energy}(z)$ **decreases** accordingly all the travel times that the driver was experiencing on every edges in path z : $\sum_{e \in z} T_e(x)$

It is like that the driver, abandoning path z for the new path z' , releases a potential energy that is equal to: $\sum_{e \in z} T_e(x)$

continues to the next page

By the same reasoning, for every edge e' in the new path z' , before the new driver adopts it, we have this potential energy:

$$\text{Energy}(e') = T_{e'}(1) + T_{e'}(2) + \dots + T_{e'}(x-1)$$

When one of the new driver joins it increases to:

$$= T_{e'}(1) + T_{e'}(2) + \dots + T_{e'}(x-1) + T_{e'}(x)$$

Summing up, Energy(z') **increases** accordingly all the travel times that the new driver is experiencing on every edges in path z' : $\sum_{e' \in z'} T_{e'}(x)$

\Rightarrow The *net change* in potential energy is simply the driver new travel time minus their old travel time

$$\Delta E = \sum_{e' \in z'} T_{e'}(x) - \sum_{e \in z} T_e(x)$$

$\Delta(E)$ must be **negative**, because driver must have an incentive to change path (the new strategy must be a best response) \Rightarrow **the potential energy strictly decreases throughout the process**

□

58.8 Comparing Equilibrium traffic to the Social Optimum

We proved that an equilibrium traffic pattern always exists, and now we want to compare the travel time experienced in an equilibrium traffic pattern to the time of a social optimum pattern. By definition, if we have a traffic pattern in social optimum, that means that the social cost is lower compared to all the other possible travel times that can be computed for all the other patterns, since it is social optimum by definition.

We hence need a relationship between the potential energy of an edge and the total travel time of all the drivers crossing the edge and then we can sum up these quantities for all the edges in the traffic patterns and compare travel times at equilibrium and at social optimum.

58.9 Potential Energy and Total Travel Time (TTT) for an Edge

$$\text{Energy}(e) = T_e(1) + T_e(2) + \dots + T_e(x)$$

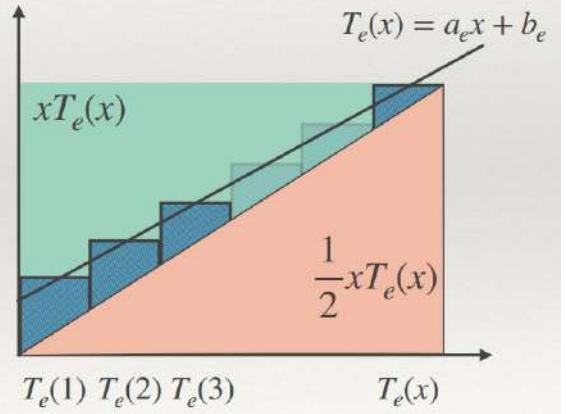
$$\text{TTT}(e) = xT_e(x) \Rightarrow \text{Energy}(e) \leq \text{TTT}(e)$$

$$= \underbrace{T_e(x) + T_e(x) + \dots + T_e(x)}_{x \text{ times}}$$

Recall that is a linear function: $T_e(x) = a_e x + b_e$

Geometrically, we have that: $\frac{1}{2}xT_e(x) \leq \text{Energy}(e)$

$$\Rightarrow \text{Energy}(e) \geq \frac{1}{2}\text{TTT}(e)$$



We have that $\text{Energy}(e) \leq \text{TTT}(e)$ and $\text{Energy}(e) \geq \frac{1}{2}\text{TTT}(e)$

$$\Rightarrow \frac{1}{2}\text{TTT}(e) \leq \text{Energy}(e) \leq \text{TTT}(e)$$

Moreover, if z is a traffic pattern, recall that: $\text{Energy}(z) = \sum_{e \in z} \text{Energy}(e)$

Recall also that the **social cost** of traffic pattern z is the sum of the travel times incurred by all drivers when they use this traffic pattern: $\text{SC}(z) = \sum_{e \in z} \text{TTT}(e)$

Finally, recall that the potential energy decreases as best-response dynamics moves from z to z' : $\text{Energy}(z') \leq \text{Energy}(z)$

58.10 Total Travel Time at Equilibrium and at Social Optimality

If z is the traffic pattern at social optimality, and z' is the traffic pattern at the end of the best-response dynamics (i.e., at equilibrium), we have that:

$$\text{Energy}(z') \leq \text{Energy}(z)$$

Moreover, we have:

$$\text{SC}(z') = \sum_{e' \in z'} \text{TTT}(e') \leq \sum_{e' \in z'} 2 \cdot \text{Energy}(e') = 2 \cdot \sum_{e' \in z'} \text{Energy}(e') = 2 \cdot \text{Energy}(z')$$

$$\text{and } \text{Energy}(z) = \sum_{e \in z} \text{Energy}(e) \leq \sum_{e \in z} \text{TTT}(e) = \text{SC}(z)$$

$$\text{then } \text{SC}(z') \leq 2 \cdot \text{Energy}(z') \leq 2 \cdot \text{Energy}(z) \leq 2 \cdot \text{SC}(z)$$

□

58.11 Conclusions of Traffic Game

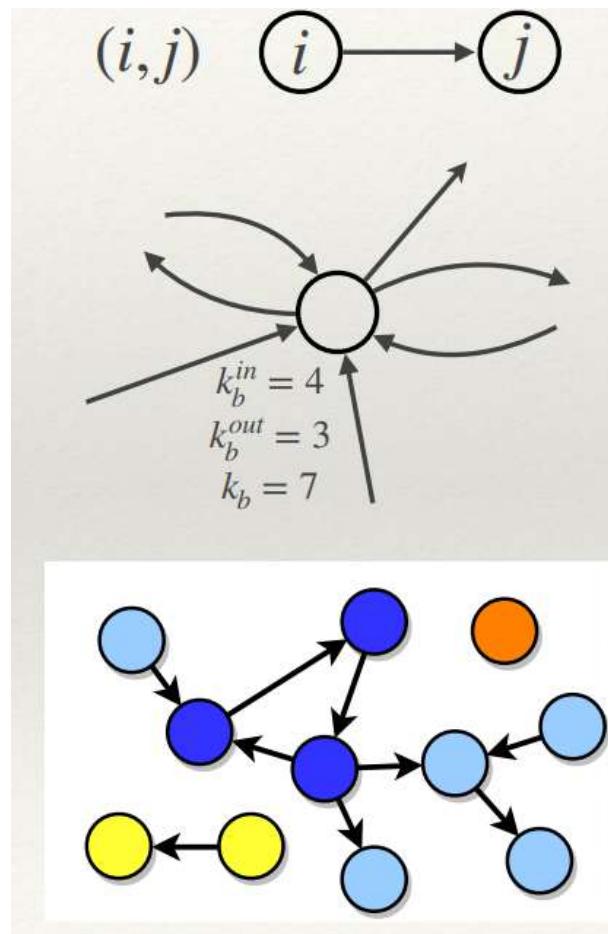
We have found that:

- In the traffic game we can always find a traffic pattern at equilibrium
- That the social cost of the traffic pattern at equilibrium is at most twice the socially optimal cost, we found a bound of the worst possible scenario.

59 DIRECTED NETWORKS

Many real-world networks are directed and/or weighted such as:

- **Food webs:** (directed) links go from prey to predator and (weights) represent amount of prey/energy consumed
- **Web and Wikipedia:** links go from source to target page and may have weights representing clicks
- **Social Media (as Twitter):** links may be weighted by numbers of interactions, such as retweets, mentions, likes, comments, etc.
- **Citation Networks:** citations from one scientific paper to another
- **Communication Networks** (including Who-Talks-to-Whom networks)



In directed networks we have a direct link from a **source to a target**, the node degree is split into **in-degree and out-degree** and the paths must respect and follow the link directions, and we know that the paths can lead to a node but we can't come back to our previous node from that node and this makes the problem of finding components and connected components a bit more complicated, but we can use strong and weak connectivity to find these components

59.1 Examples of Directed Networks

Information Networks are an example of directed networks where the **units** are the pieces of the network and the links are the relationships between these units. Information networks can be citation network, the web, wikipedia etc..

In a **citation network**, nodes are research articles and links represent citations of references where links tend to point strictly backward in time and hence there is a flow of time that returns a sense of flow from present to past

Bibliographic Coupling is similar to a citation network where we have semantic similarity measures such as **citation and co-reference**. **Co-citation networks** have a number of shared predecessors (cited by same papers), while **Co-reference networks** have a shared number of shared successors (citing same papers)

60 THE WEB AND ITS STRUCTURE

The World Wide Web is a way to exchange documents via the Internet, where the web pages are public pages and the browser allows us to connect and access the internet. Links between pages are called **hyperlinks** that allow us to go from a page to another, while the hypertext is the organization of the information using a network metaphor. The **hypertext** is not a simple graph like a path (for example in a book we know that we need to follow the order of the pages), but we have a graph that connects the pages in some way that is organized specifically.

The links that are formed in some of the directed networks such as the WWW or even Wikipedia, follow a formation and a linking that functions similarly to how our brain creates links between thoughts and stuff

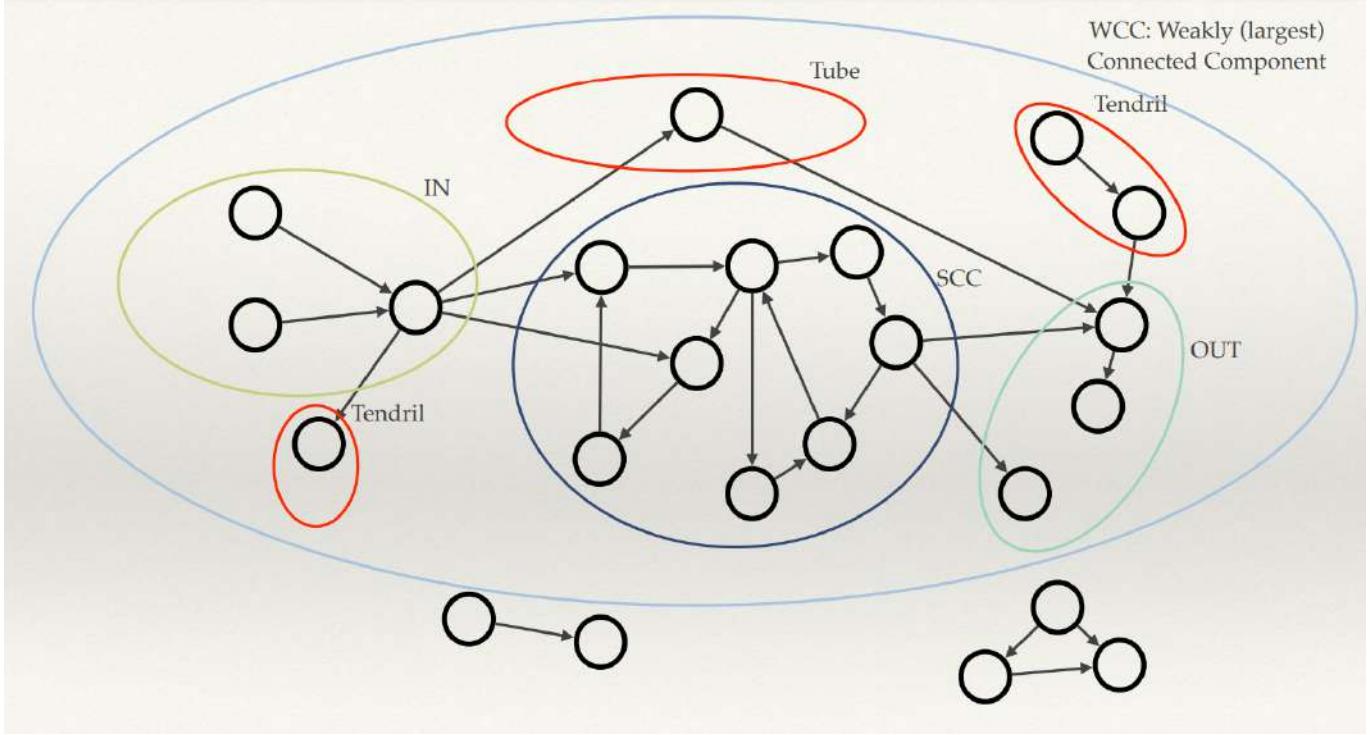
60.1 The Web as a Directed Graph

We need to formalize the web in terms of measures such as shortest path, connected components, degree distribution etc.. The informative power of the underlying structure of the graph structure of the web is quite big since we can do things such as:

- understand better the logical relationships expressed by its links if we analyze the structure of the web
- break its structure into smaller units
- identify important pages as a step in organizing results of the web, which is the most important function offered by the web since it is massively important to identify important pages

60.2 The Structure of The Web

Broder used Altavista to build a map of the Web dividing the graph into few pieces. They then crawled the Web from large companies sites using a BFS algorithm and as a result they would get a 'piece of the web'. At the end of the process they found a largest connected component (or Giant component), containing a significant fraction of all the nodes.

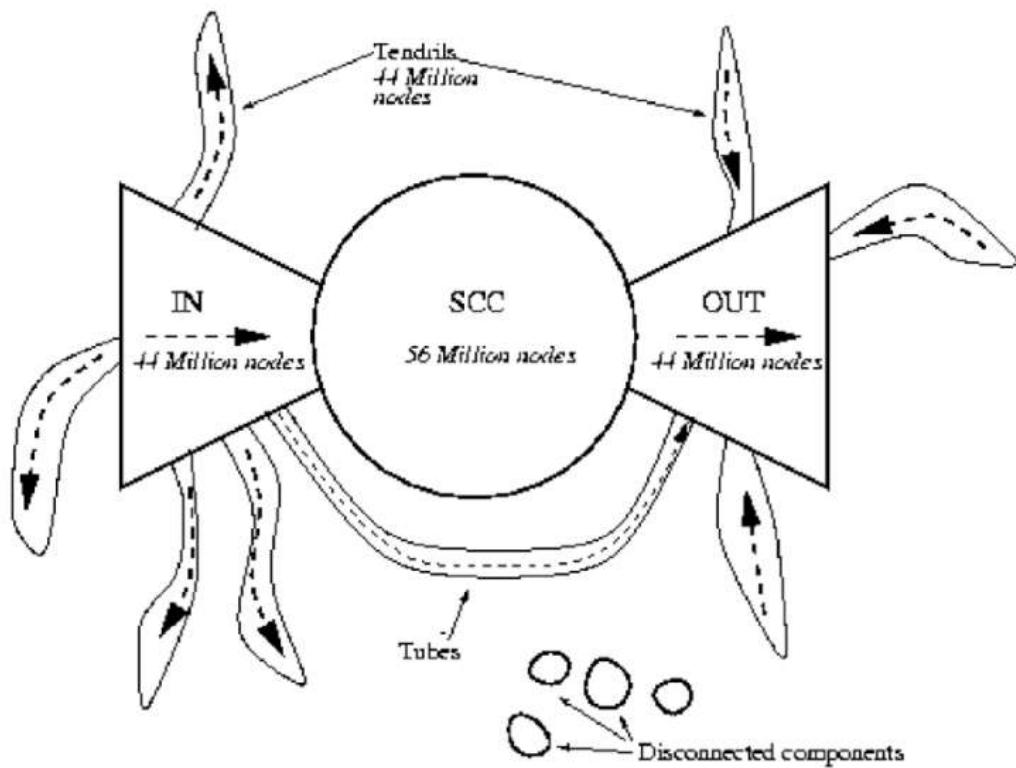


The results showed that there was a **WCC (Weakly Connected Component)**, weak because we can have a path from a source to the target but not the way back, and this weakly connected component contained a significant fraction of the nodes that composed the web at the time. Outside of the WCC we can have small components that the WCC can't reach. At the core of the WCC we have a **SCC (Strongly Connected Component)** where every pair of sources and targets have a path from the source to the target and viceversa. Inside of this SCC we find **core nodes** that can be reached by many others, and we need to go through this nodes to reach the nodes on the other side of the graph.

The nodes of the **IN Components** have a path that leads to a node in the SCC even if there is no route back, and the **OUT Components** are the nodes that have a path to them starting from the nodes of the SCC.

The **Tubes** are nodes that are part of the path from an IN component to an OUT component that doesn't go through the SCC. While the **Tendrils** are nodes reachable from IN components or nodes that reach OUT components.

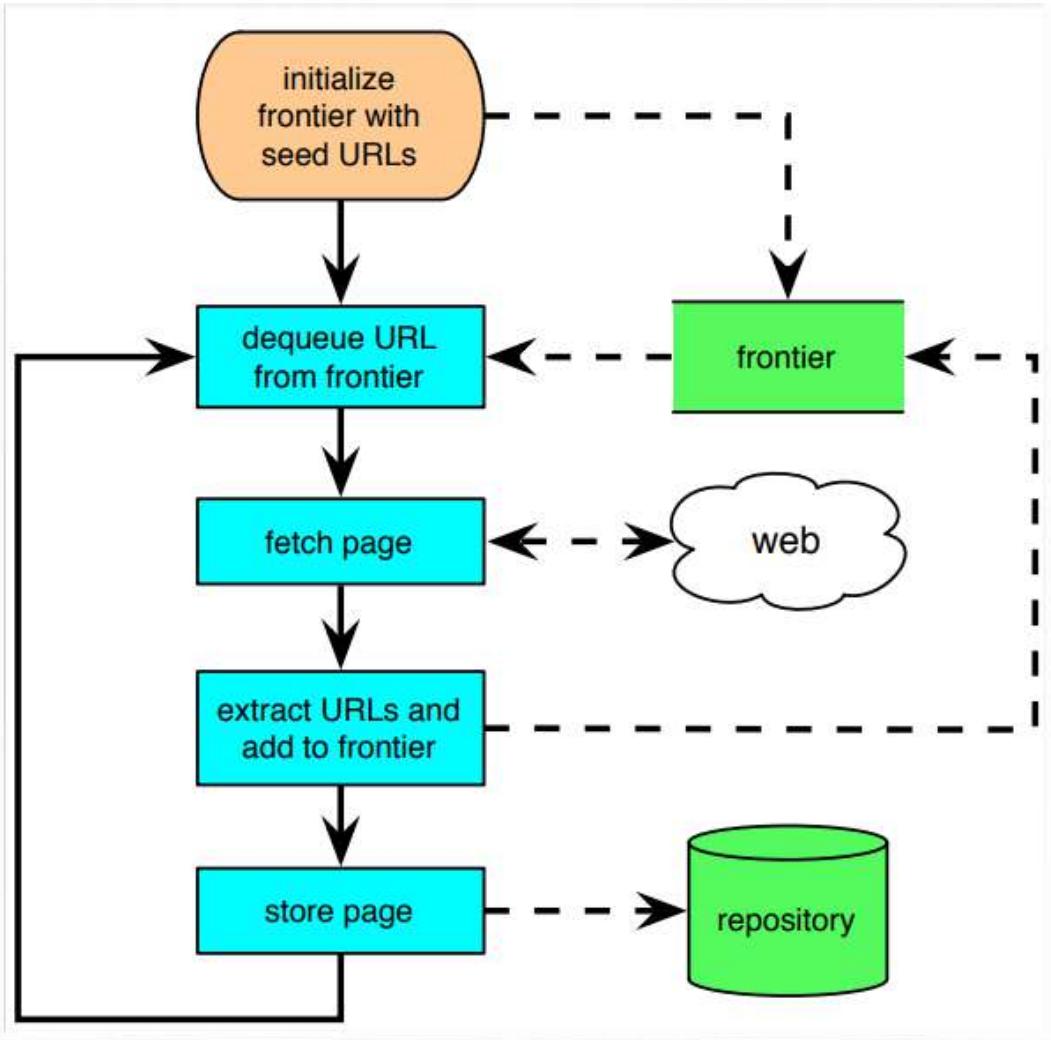
60.3 The Bow Tie Structure of The Web



Broder even found out the sizes of these components, where the SCC is the biggest in size, but even the in, out components and the tendrils have quite big sizes. He determined that if you follow a tendril you will never be able to reach the SCC, and this has a massive impact on the design of **web crawlers**.

60.4 Web Crawlers

Web crawlers are programs that automatically download web pages and retrieves them so that we can work on them and organize them accordingly, the primary application of crawlers are search engines. Crawlers collect information, scattered across billions of pages served by millions of servers around the globe, to a central location where it can be analyzed and mined. A search engine takes the information collected by a crawler and creates an index to efficiently retrieve pages that contain keywords and phrases queried by users. The Web changes rapidly, so search engines use crawlers weekly to stay fresh information as pages and links are added, deleted, moved, and updated.

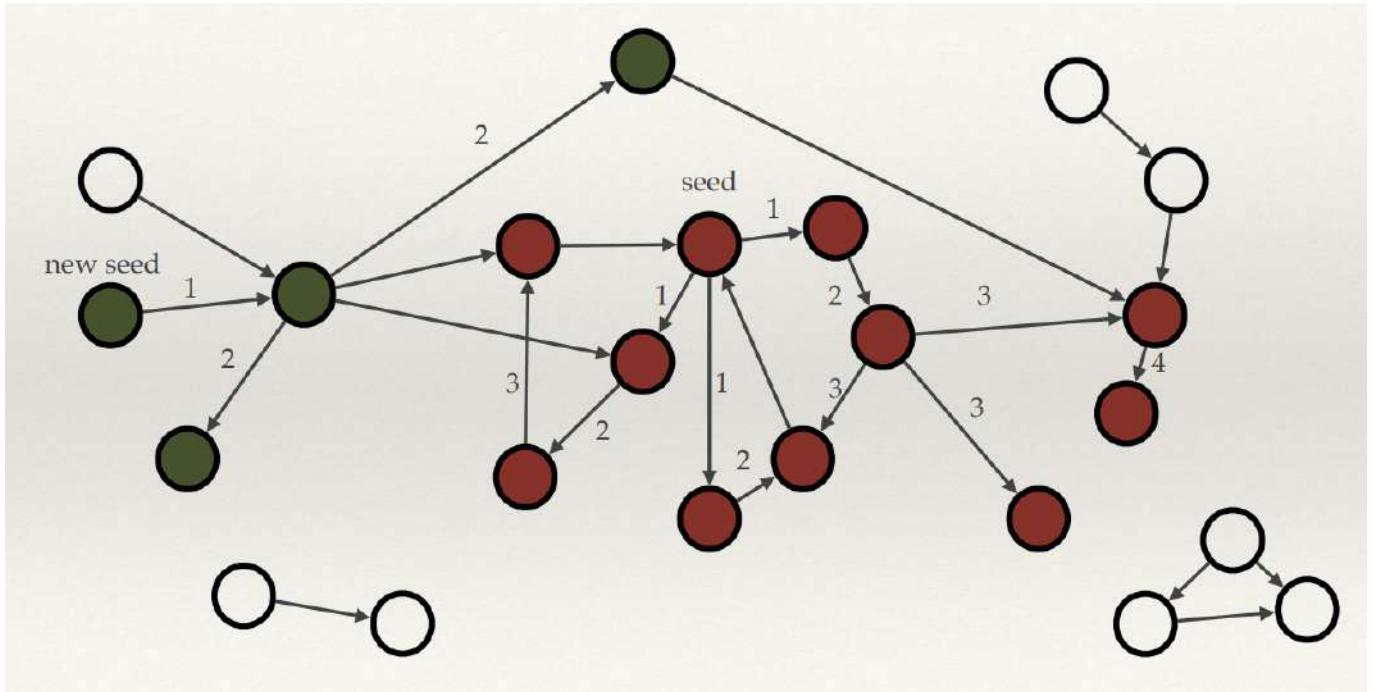


The idea behind a crawler is quite simple, we run a breadth-first traversal of the Web graph, as in the BFS algorithm:

- Start from a set of high-quality seed pages
- The queue of unvisited URLs is called frontier
- Fetch pages dequeued from the frontier, extract links and add them to the frontier and repeat the BFS algorithm after we have saved the pages in a repository

We can run into a lot of problems due to scalability, page revisit scheduling since we need to always know when a page has been updated, dealings with servers all around the world etc..

60.5 Which Page is a Good Seed Page?

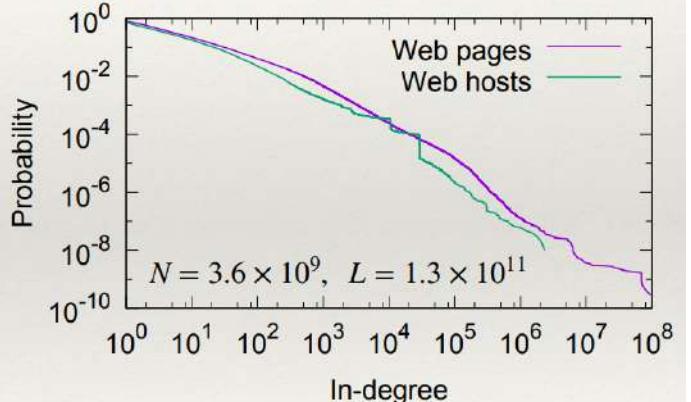


If we start with a seed in the SCC we know that we will cover all the **strongly connected components and the out components**, but if we want to get the other nodes we need to consider and start with a seed in the **IN components**. Starting with a seed in the SCC or IN components brings a risk to miss and skip some tendrils and some connected components.

Understanding in which components every node belongs is very important. Of course we need more than one seed page, but at least one of these seeds must be selected in the IN components. We also need to accept that even 50 percent of the pages could miss, and we need to plan our engineering accordingly, but reaching all the pages is virtually impossible.

- ❖ We focus on **in-degree** distribution, which arises from collective behavior of authors
 - ❖ Heavy tail with huge hubs:
- $$\langle k_{in} \rangle \approx 10, \quad \kappa \approx 40$$
- ❖ Out-degree distribution is less meaningful as it depends on individual content providers and can be gamed
 - ❖ Eg, **link farms** for **spamdexing**
 - ❖ Short paths in core:

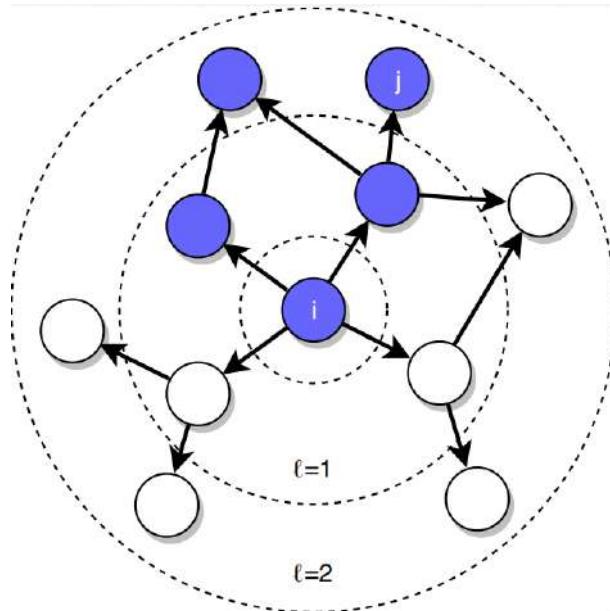
$$N \approx 1.8 \times 10^9, \quad \langle \ell \rangle \approx 13$$



Hubs leads to the **ultra small world** phenomenon!

In web degree distribution we focus on in-degree distribution, which arises from collective behavior of authors of pages and not on out-degree distribution since it is less meaningful as it depends on individual content providers and can be used for link farming and other bad uses.

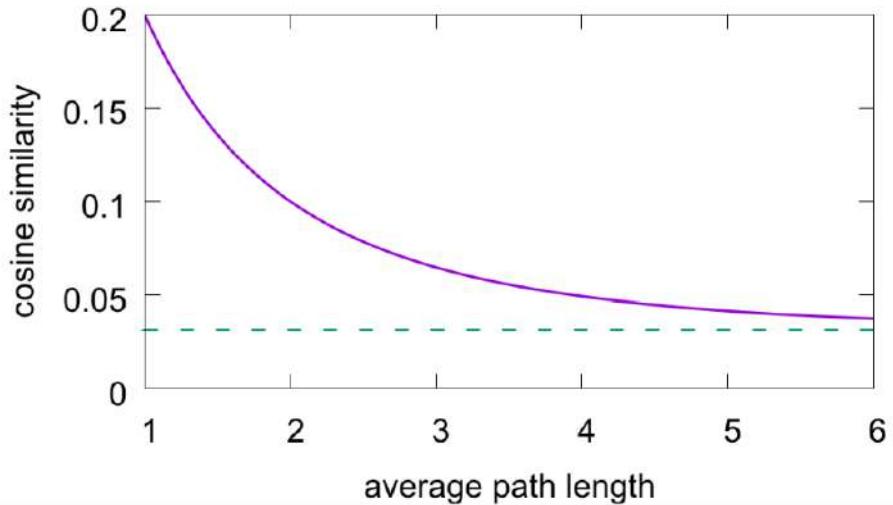
60.6 Topical Locality



Is a property that refers to the Homophily, which happens when two nodes are linked with each other and they are also very similar. In the web **homophily** is defined as pages about related topics that tend to have links between each other. If we start from a certain page we can find some pages with similar topics, and more we move from the starting page, the lower is the chance to find similar topics.

Links between pages are not random, and we can think of **similarity based on textual content or categorical topics**.

60.7 Topic Locality and Topic Drift



If we have a defined similarity between pages in function of the distance between two pages in the graph, we can determine that the similarity between the pages decreases significantly while the distance between pairs of nodes decreases, and hence we have a signal of **Topic Locality**.

While we browse away from a starting page, the similarity decreases and we have a **Topic Drift**. This is something like a stream of consciousness from the idea of Jaymes Joyce.

Topical locality makes Web browsing possible as we stay on topic, and crawlers use BFS because of topical locality to target and get the good quality pages and sites that are linked to other good quality sites.

60.8 Cosine Similarity

Cosine Similarity is often applied to textual content of pages, we can represent every page of the web with a vector of values (d), where the values could be the frequency of certain given words (W_d) in that document or page.

$$\vec{d} = \{w_{d,1}, \dots, w_{d,n_t}\}$$

We could have a certain document that could be a file, a tweet, a web page etc.. and we have another document and we want a vectorial representation of these two documents and then we want to understand and compute the similarities between these two vectors.

Information retrieval is an important technique to retrieve useful information in documents.

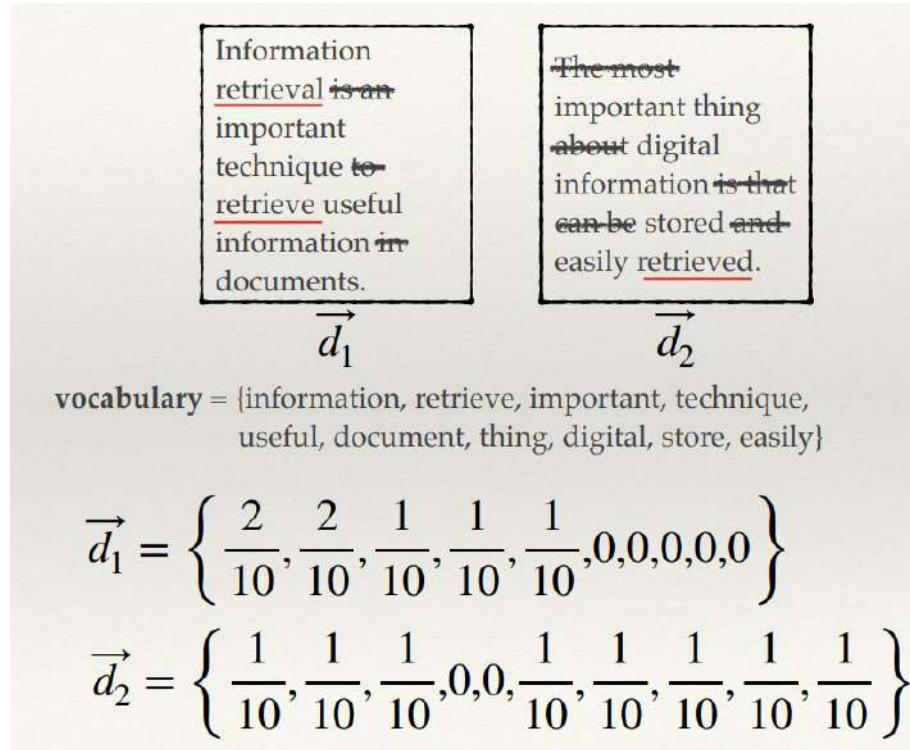
$$\vec{d}_1 ?$$

The most important thing about digital information is that can be stored and easily retrieved.

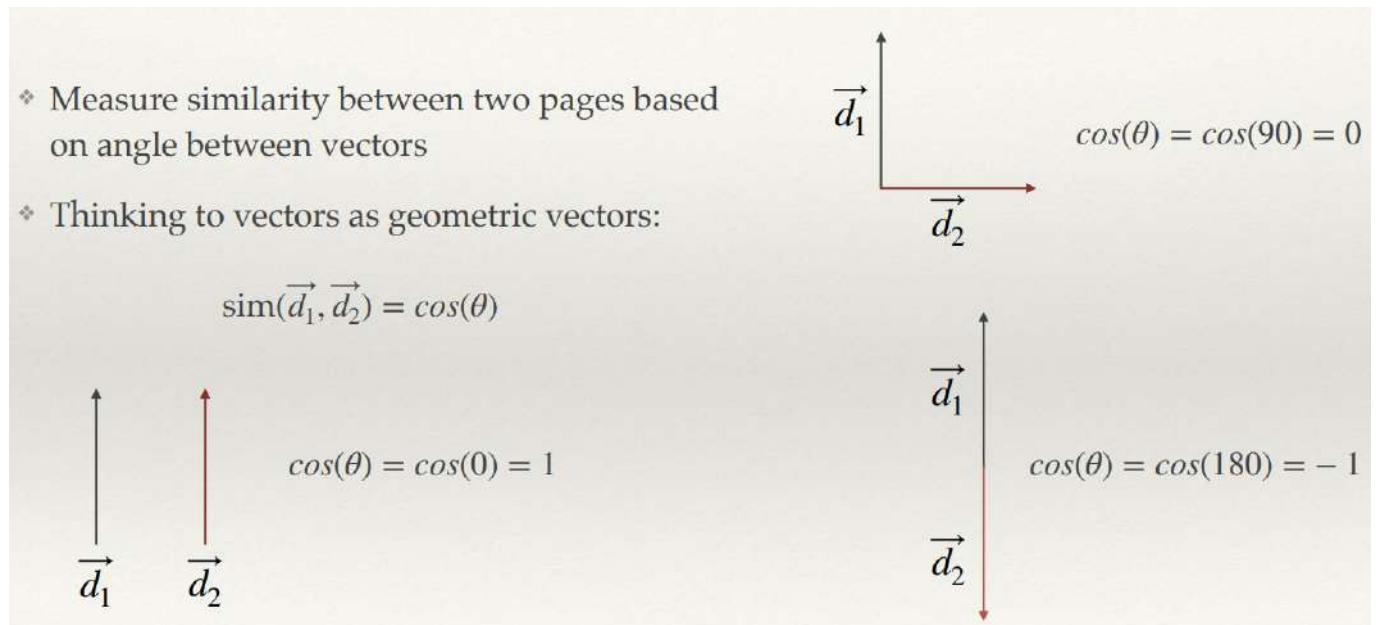
$$\vec{d}_2 ?$$

We have one dimension for each term in the vocabulary and hence we have a high-dimensional space problem, and to try and mitigate this problem we can remove or give lower weights to terms that are very common and therefore not meaningful, or we could use deep learning techniques to reduce the size of the problem since we could find words in the document that are not useful to compute the similarity.

If we want to reduce dimensionality we need to remove **stop words**, we need to find **common roots of certain words** and then we create a **vocabulary** of the important words that we can use and at the end of the process we create a vector that characterized each document.



We now want to measure the similarity between these two pages using their vectors. If we think about the vectors in a geometrical way we can consider the angle between these two vectors and try to determine how similar they are.



We can do the same thing but in an algebrical way using the Euclidian dot product, where the product of two vectors is the product of the norm of the two vectors and the cos0.

$$sim(\vec{d}_1, \vec{d}_2) = \cos\theta$$

Euclidean dot product

$$\vec{d}_1 \cdot \vec{d}_2 = \|\vec{d}_1\| \cdot \|\vec{d}_2\| \cdot \cos\theta$$

$$= \frac{\vec{d}_1 \cdot \vec{d}_2}{\|\vec{d}_1\| \cdot \|\vec{d}_2\|}$$

$$= \frac{\sum_t w_{d_1,t} \cdot w_{d_2,t}}{\sum_t w_{d_1,t}^2 \cdot \sum_t w_{d_2,t}^2} \quad \in [-1,1] \quad \begin{cases} -1 & \text{exactly opposite} \\ 0 & \text{orthogonal} \\ 1 & \text{exactly the same} \end{cases}$$

We can write the cosine, and hence the similarity of these two vectos can be defined as the eucliudean dot product divided by the product of the norms of the two vectors. If we calculate the products, which is the sum of the values in the two vectors divided by their norms.

If the result is -1 the vectors are totally opposite, if the result is 0 they are orthogonal and can't compare, but if their similarity go to 1 we have that the vectors are exactly the same.

60.9 Pearson's Similarity

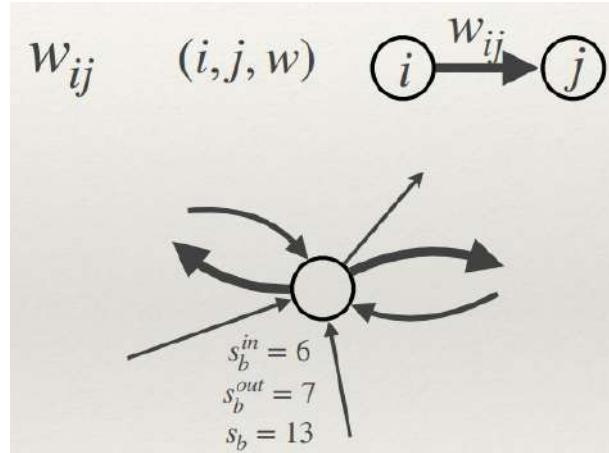
Is exactly the same of the cosine similarity but we subtract the vector means

- ❖ As the cosine similarity, but the attribute vectors are normalized by subtracting the vector means:

$$r(\vec{d}_1, \vec{d}_2) = \frac{\sum_t (w_{d_1,t} - \bar{w}_{d_1}) \cdot (w_{d_2,t} - \bar{w}_{d_2})}{\sum_t (w_{d_1,t} - \bar{w}_{d_1})^2 \cdot \sum_t (w_{d_2,t} - \bar{w}_{d_2})^2}$$

61 WEIGHTED NETWORKS

A lot of real life networks have link weights such as twitter retweets, the emails, the data exchanged between routers, but even facebook friendships could be weighted like the tags in a post, likes etc.. even if facebook is undirected and unweighted.



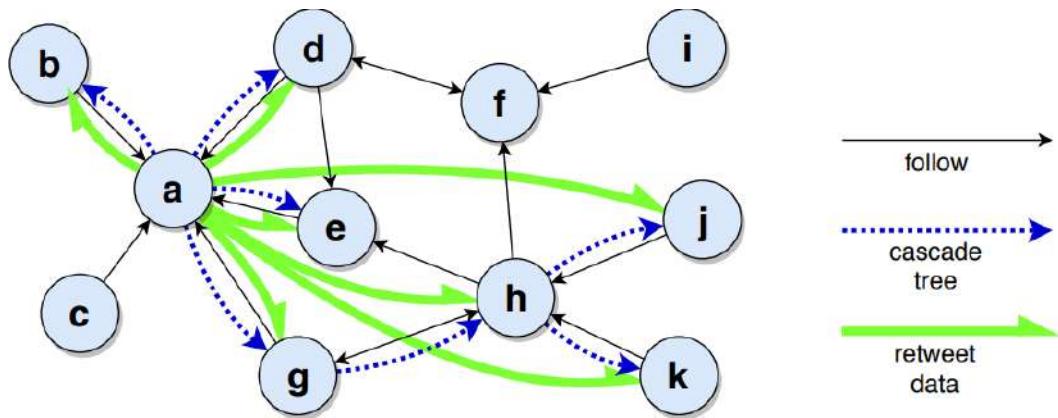
Degree, in-degree and out-degree are extended to **strength**, **in-strength**, and **out-strength**

61.1 Information Diffusion Networks

We can have things such as memes that are transmissible units of information, such as ideas, behaviors, news links, hashtags, and, yes, also images with captions (image macros) that can be used to understand certain behaviors and ideas. The definition of meme is due to Richard Dawkins, that stated it as an analogy to genes transmitted from parent to offspring and like genes, memes can mutate and be transmitted to many other people that are close to us.

Tweets and posts can carry high amounts of memes and we can track, map, and analyze the spread of memes on Twitter using certain tools.

61.2 Retweet Networks



In the data, each retweet cascade network is a star which means that all retweets point to original tweet. The actual cascade tree is difficult to reconstruct, but we can make some guesses based on the follower network and timestamps.

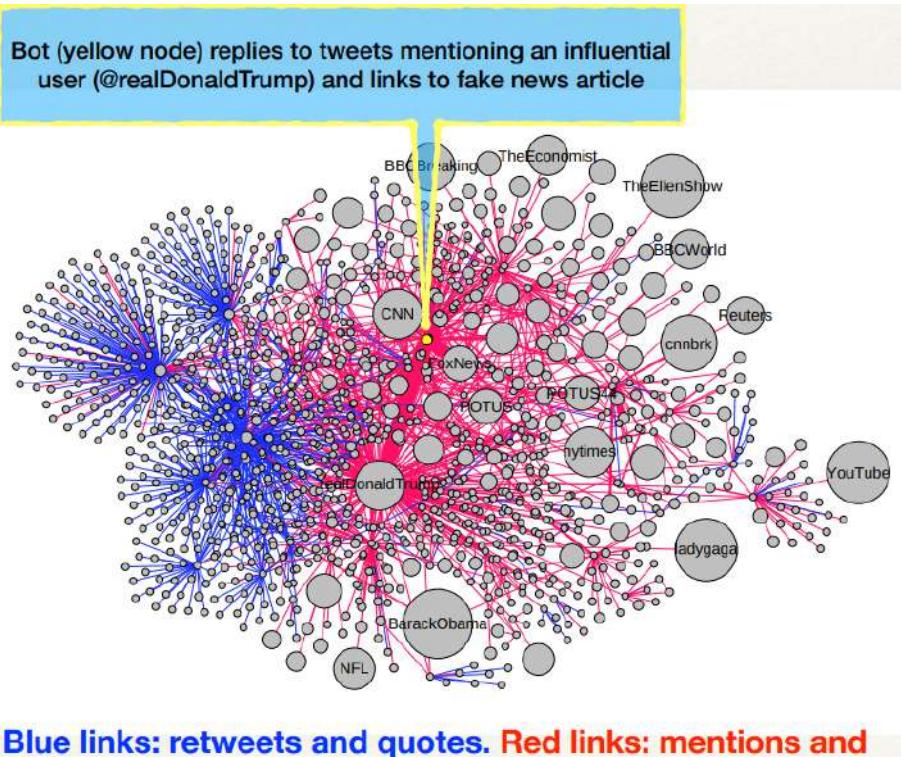
61.3 Virality

There are multiple ways to measure the virality of a meme:

- Number of users exposed
- Depth of diffusion tree
- Fraction of users who retweet to users who are exposed

The bad thing about this is that misinformation is often more viral than actual news reports

61.4 Influence



Certain bots can be influential for spamming and sharing fake news. In the retweet networks of the presidential campaigns it was found that a bot was spamming fake news and donald trump retweeted this fake news and it spread and influenced a lot of people.

Social Bots are account controlled by software and have malicious intents and manipulate the diffusion of fake news in the networks

Multiple ways to measure the influence of an account:

- Number of followers (in-degree in follower network)
- Number of users exposed (out-degree in retweet network)
- Number of retweets (out-strength in retweet network)
- Fraction of retweets to followers
- Social bots can target influential accounts hoping for retweet

61.5 Co-occurrence Networks

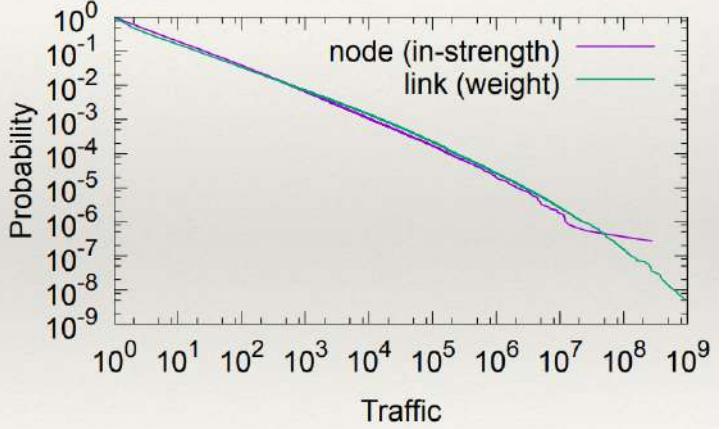
We can build this network from real data. In this networks weight can originate from relationships between more than one type of entity, where nodes in a bipartite network **co-occur** if they share common neighbors. We have seen that maybe people that visit frequently a certain site or place tend to form friendships and link with each other.

We have hence a **Projection**, that keeps nodes from either group of bipartite network and add edges among co-occurring nodes

Co-occurrence networks can be used for different ways such as **folksonomy**, that are sets of triples (user, resource, tags) where an user assigns a tag to a resource, and this forms a **tripartite network** where each link connects three nodes from distinct sets. Resources can be photos, hashtags etc.. and we can observe how often there are connections between web pages and users, between sites and the type of information they talk about etc..

61.6 Weight Heterogeneity

- ❖ Weights can span many orders of magnitude
- ❖ Example: Web traffic
 - ❖ Click data collected at IU
 - ❖ ~100k anonymous users
 - ❖ ~1B clicks
 - ❖ $N \approx 4M$ sites
 - ❖ $L \approx 11M$ weighted directed links
- ❖ Both in-strength (s_{in}) and weight (w) distributions have very heavy tails



As we compute the degree distributions and we plot them we can talk about weight distributions. We have very few hubs at the tail that have an high weight, but the majority of the nodes on the network have a very low weight in the graph.

61.7 Link Filtering

We might have many links with very small weights, making a weighted network too dense and complex to easily visualize or analyze. If low-weight links represent low-importance connections we could filter them out to analyze a sparser network and the easiest and most efficient approach would be to delete and remove the links with weights below a certain **threshold of importance**.

The problem of implementing this is that we would remove a consistent part of the network, and if we have a constant threshold what we are doing is just changing the shape of the network itself, but this approach works well enough in many cases.

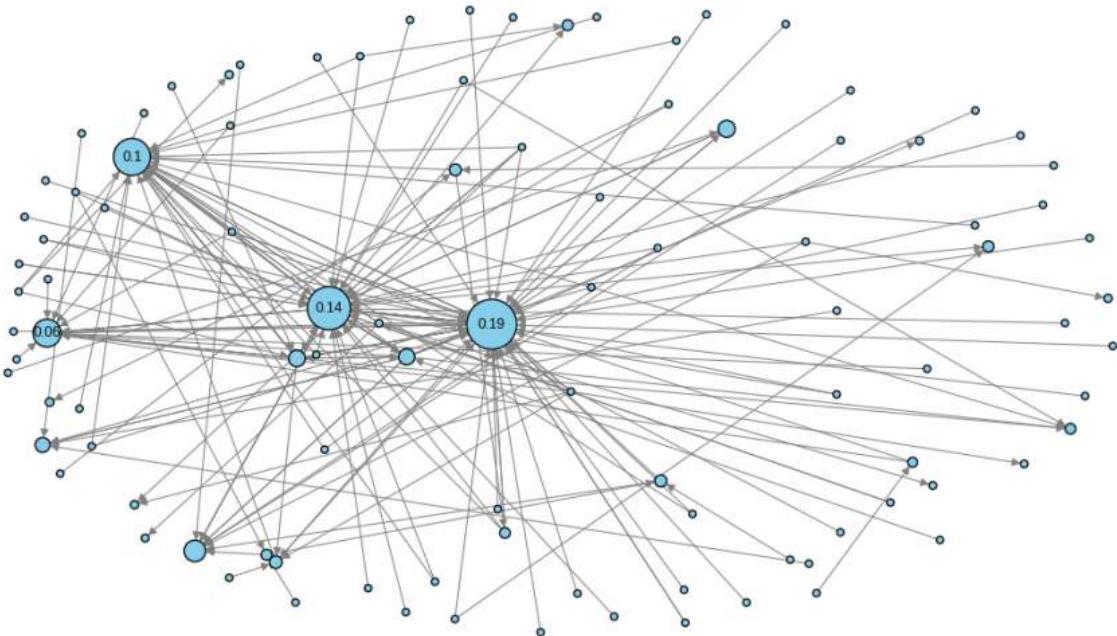
But how to set the threshold to filter links?

We know that due to weight heterogeneity, a lot of information is lost by setting a threshold such that enough links are removed and so we want to set a **Backbone**, which means to set a different threshold for each node, so that we remove links with weights that are not significant compared to the other links of that node, this is somewhat of a significance threshold that could preserve the structural properties of the network.

We keep the link if P_{ij} is lower than the threshold we imposed before.

$$P_{ij} = \left(1 - \frac{w_{ij}}{s_i} \right)^{k_i-1} < \alpha$$

62 SEARCHING THE WEB



PageRank and HITS are the **centrality measures** for nodes in directed networks. Sergey Brin and Larry Page introduced PageRank in 1998 as a key ingredient of Google while Jon Kleinberg introduced HITS in 1999. Both these methods are based on eigenvector centrality and designed for web information retrieval.

62.1 The Problem behind Searching The Web

Search engines are massively important to reduce the information overload problem related to a given topic in the internet, and so we ask the engines to find pages related to a given topic and also to rank these web pages from the most important to the least important.

Information retrieval is a computer science subfield that introduced automated strategies to search in libraries, scientific papers, repositories etc.. in response to keywords given in input to the search engine, and the engine transforms these inputs to queries that need to give the desired result in output. The keywords used to search on the engines are quite inexpressive and we have a lot of diversity on given topics and hence we need a ranking to get the best results that are needed to us.

Pages are dynamic and they always change their format and content and so we need to understand and filter what information is important and what isn't, and the structure of the web can help us to define such filters. An idea could be the one of counting the words of a document and rank with an heavier weight the pages that contain the information that we want to retrieve much more then other pages, but we can do much better since if we want to exploit the structure of the web to our advantage we need to focus on more then just the content of a web page.

63 HITS – LINK ANALYSIS USING HUBS AND AUTHORITIES

HITS is based on finding nodes of two different categories called **Hubs or Authorities**. The idea is to adapt the notion of a node being an hub in a directed network such that we can differentiate the behavior of a node that is a proxy to other important nodes, or a node that has a lot of important paths pointing to it.

We want to use not only the information contained in the page itself with word counting, but also the information contained between pages as well. We select the certain documents regarding a given topic and we count the **in links**, which are a measure of the authority of a page on such a topic, it is an implicit endorsement from the community of web pages' authors. If we are creating an hyper link to another page, we are saying that this page is important regarding that given topic.

63.1 Finding Lists and List Value

If we query a certain word we wouldn't really be able to differentiate the results that are proposed to us based on who gave that result, so we could find intuitively correct authority values by just counting the in links. A node could be considered a correct authority if it has many links going inwards to him. In this situation we could have as a return to our query some nodes that are **lists**, which are pages that provide a lot of outlinks to other pages. This allows us to divide roughly the pages returned from our query in pages that have authority since they have a lot of in links in them, and other pages that are lists or hubs that are more likely to provide links to other pages instead of having the needed information in them.

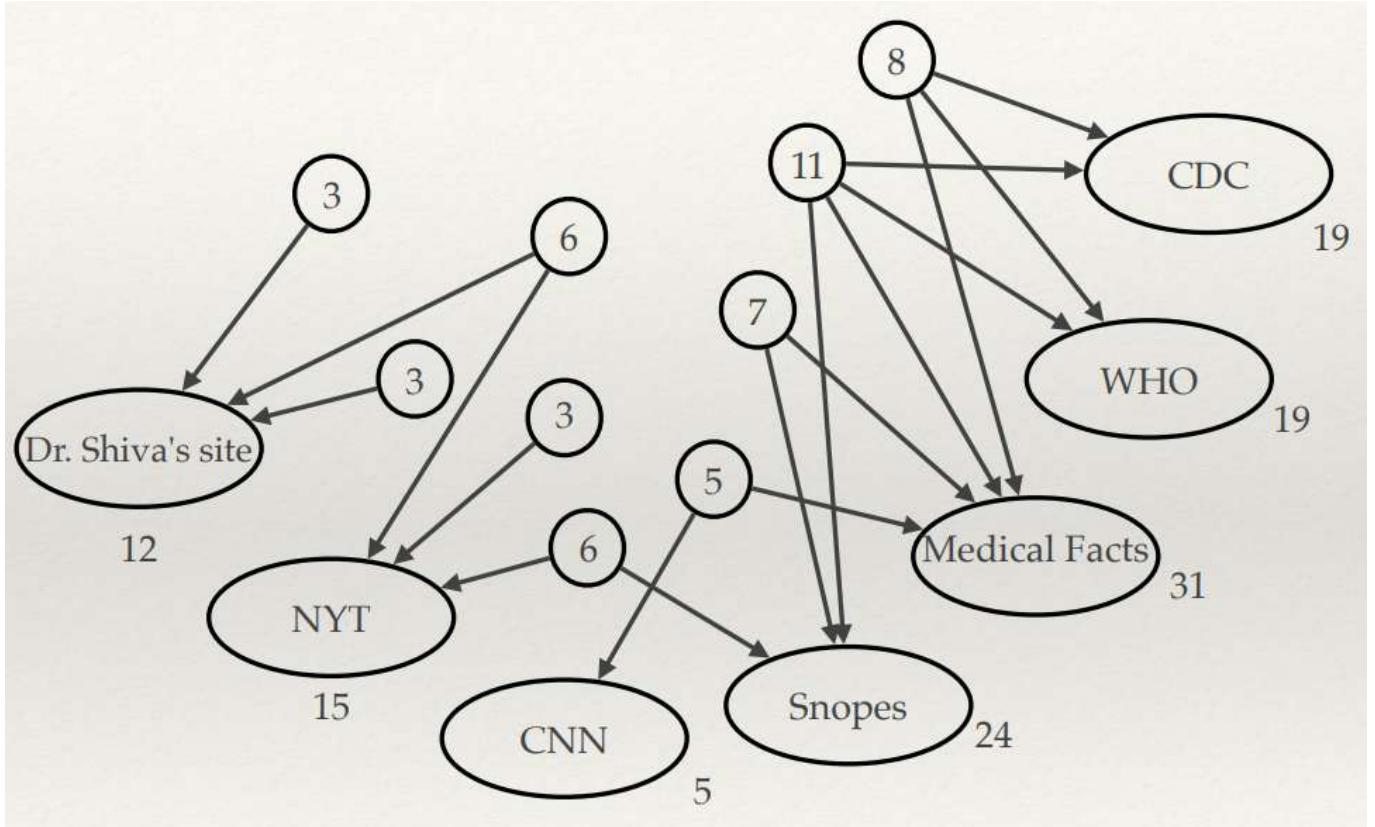
We then can assign values to the authority pages based on the number of incoming links that they have. We want to calculate the hub values and we follow a certain assumption that list pages have a better sense for knowing where the good results are, so the hubs/lists are very important since they allow us to understand where the good results are lying.

63.2 Calculating Hubs and Authorities

To calculate hub values we just move the value of the authority to which an hub is connected to the value of the hub itself, if an hub is connected to more authorities we sum all the values of his authorities. Now that we have recalculated the hub values, the importance of receiving a link from a hub whose value is 3 and one whose value is 6 is much more different, so we do the same process we did with the hubs but for the authorities, where their value becomes the sum of the values of the ingoing hubs to the authorities themselves.

We could sum up the algorithm in this steps:

- step 0: every node has both a hub and an authority value and we initialize them all to 1
- step k (**Update Rules**):
 - authorities are updated after k-1's hub values
 - hubs are updated on new authority values
- normalize hub and authority values



Normalization is important because if we repeat this iterative method it will never stop and we need them to stop when we reach a convergence, and the normalized values converge when i tends to infinity. We can also say that initial values are not important for stabilization since limiting values for hubs and authorities are **properties of the links structure**, so they are dependent on the structure of the network and not only on the initialization values.

64 SPECTRAL ANALYSIS OF HITS

In order to prove the termination of the Hub and Authorities calculation algorithm, the stabilization and the fact that the initial values are not important for reaching limiting values we will perform a spectral analysis of HITS.

We need to analyze the methods to compute hubs and authorities values, and we will be able to prove that limiting values are coordinates in eigenvectors for given eigenvalues in matrices derived from our graphs, and when we use eigenvectors and eigenvalues to study the structure of the network we say that we are performing a spectral analysis.

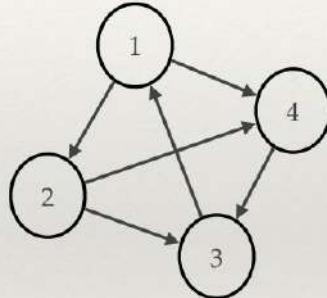
We will use an adjacency matrix M to begin

def.: Adjacency Matrix (*)

nodes: $1, \dots, n$

$M : n \times n$

$$M_{ij} = \begin{cases} 1, & \text{if } (i, j) \text{ is an edge} \\ 0, & \text{otherwise} \end{cases}$$



$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

(*) not necessarily efficient for computational representations (for practical use, consider adjacency lists or edge lists instead)

64.1 Update Rules

\mathbf{h}, \mathbf{a} : n-dimensional vectors (resp., hub and authorities values)

Hub Update Rule

$$h_i \leftarrow \sum_{j=1}^n M_{ij}a_j$$

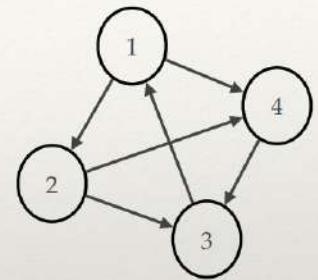
$$= M_{i1}a_1 + M_{i2}a_2 + \dots + M_{in}a_n$$

$$\mathbf{h} \leftarrow \mathbf{M} \cdot \mathbf{a}$$

Authority Update Rule

$$a_i \leftarrow \sum_{j=1}^n M_{ji}h_j$$

$$\mathbf{a} \leftarrow \mathbf{M}^T \cdot \mathbf{h}$$



$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 1 \\ 2 \end{bmatrix}$$

\mathbf{M} \mathbf{a} \mathbf{h}

\mathbf{h}, \mathbf{a} are two vectors with n dimensions since every node can have an hub value and an authority value, and since we have n nodes these two vectors must have n dimensions.

The **Hub Update Rule** can be defined with a sum, if we want to update the node i we need to check how many incoming edges we have from i to j (M_{ij}) where j is a node whose authority value is a_j . We could also use the vectorial representation by saying that vector \mathbf{h} is updated by the product of the matrix \mathbf{M} with the vector \mathbf{a} .

The **Authority Update Rule** is kinda the same since the authority value of node i (a_i) is just the sum for all the edges that are connecting j to i, so all the incoming links, multiplied by the hub value of \mathbf{h} . Using the vectorial representation \mathbf{a} is given by the product of transpost \mathbf{M} with \mathbf{h} .

64.2 Understanding the k-step Hub Authority Computation

init: $\mathbf{h}^{<0>} = \underbrace{(1, 1, \dots, 1)}_n$

after k applications of update rules: $\mathbf{a}^{<k>}, \mathbf{h}^{<k>}$

1st: $\mathbf{a}^{<1>} = \mathbf{M}^T \mathbf{h}^{<0>}$
 $\mathbf{h}^{<1>} = \mathbf{M} \mathbf{a}^{<1>} = \mathbf{M} \mathbf{M}^T \mathbf{h}^{<0>}$

2nd: $\mathbf{a}^{<2>} = \mathbf{M}^T \mathbf{h}^{<1>} = (\mathbf{M} \mathbf{M}^T) \mathbf{M}^T \mathbf{h}^{<0>}$
 $\mathbf{h}^{<2>} = \mathbf{M} \mathbf{a}^{<2>} = (\mathbf{M} \mathbf{M}^T)(\mathbf{M} \mathbf{M}^T) \mathbf{h}^{<0>} = (\mathbf{M} \mathbf{M}^T)^2 \mathbf{h}^{<0>}$

...

k -th: $\mathbf{a}^{<k>} = (\mathbf{M} \mathbf{M}^T)^{k-1} \mathbf{M}^T \mathbf{h}^{<0>}$
 $\mathbf{h}^{<k>} = (\mathbf{M} \mathbf{M}^T)^k \mathbf{h}^{<0>}$

a, h vectors: multiplication of an initial vector $\mathbf{h}^{<0>}$ by larger and larger powers of $\mathbf{M}^T \mathbf{M}$ and $\mathbf{M} \mathbf{M}^T$

At the beginning $\mathbf{h}^{<0>}$ has a size of n and is set to all 1, so every node in the networks has an eigenvalue of 1. We want to understand after k applications of the update rules, which is the vectorial form of $\mathbf{a}^{<k>}$ and $\mathbf{h}^{<k>}$.

READ THE PROOF FROM THE SLIDES

\mathbf{a} and \mathbf{h} are vectors, but they also are the product of an initial vector $\mathbf{h}^{<0>}$ by bigger and bigger powers of $\mathbf{M}^T \mathbf{M}$ and $\mathbf{M} \mathbf{M}^T$.

64.3 Introducing Normalization – Multiplications and Eigenvectors

normalization: we can find constants c and d s.t. $\frac{\mathbf{h}^{<k>}}{c^k}$ and $\frac{\mathbf{a}^{<k>}}{d^k}$
we want to prove that they **converge** for $k \rightarrow \infty$

Focus on hub vectors:

if $\frac{\mathbf{h}^{<k>}}{c^k} = \frac{(\mathbf{M}\mathbf{M}^T)^k \cdot \mathbf{h}^{<0>}}{c^k}$ converges to a limit $\mathbf{h}^{<*>}$, then I can expect that:

$$c \cdot \mathbf{h}^{<*>} = (\mathbf{M}\mathbf{M}^T) \cdot \mathbf{h}^{<*>}$$

Hence, we need to prove that the sequence of $\frac{\mathbf{h}^{<k>}}{c^k}$ converges to the eigenvector of $\mathbf{M}\mathbf{M}^T$

At this point we introduce normalization. We can find two constants c and d such that $h^{<k>}/c^k$ and $a^{<k>}/d^k$, and we do this since we need to normalize for such constants out values at every step and hence at the denominator we have the powers of such constants.

We now want to prove that these values converge for k that tends to infinity.

If we work on hub vectors, we can see that $h^{<k>}/c^k$ is equal to the expansion we just calculated and divided by c^k could converge to a limit of $h^{<*>}$ and then we can have that c multiplied by $h^{<*>}$ is equal to $\mathbf{M}\mathbf{M}^T$ multiplied by $h^{<*>}$.

Now we need to prove that the sequence of $h^{<k>}/c^k$ converges to the eigenvector of $\mathbf{M}\mathbf{M}^T$, because the formula we just found tells us that c is an eigenvalue of $h^{<*>}$ for the matrix $\mathbf{M}\mathbf{M}^T$.

64.4 Eigenvectors and Square Matrices

- ◊ Observe that $\mathbf{M}\mathbf{M}^T$ is a symmetric matrix
- ◊ **fact 1:** "Any symmetric matrix $n \times n$ has a set of n eigenvectors $\mathbf{z}_1, \dots, \mathbf{z}_n$ that are orthogonal and all unit vectors - that is they form a basis for the space \mathbb{R}^n
 $\Rightarrow \mathbf{z}_i \cdot \mathbf{z}_j = 0$ and $\mathbf{z}_i \cdot \mathbf{z}_i = 1$
- ◊ That means that for our symmetric $\mathbf{M}\mathbf{M}^T$ we can find:
 - ◊ n mutual orthogonal eigenvectors: $\mathbf{z}_1, \dots, \mathbf{z}_n$ \leftarrow the *spectrum* of $\mathbf{M}\mathbf{M}^T$
 - ◊ n corresponding eigenvalues: c_1, \dots, c_n
 - ◊ Let's sort eigenvectors s.t. corresponding eigenvalues: $c_1 \geq c_2 \geq \dots \geq c_n$
 - ◊ Assume (for now): $c_1 > c_2$

We can observe that $\mathbf{M}\mathbf{M}^T$ is a symmetric matrix and we can tract some properties from this fact.

READ THE PROOF FROM THE SLIDES

- ◊ Let's consider $\mathbf{x} \in \mathbb{R}^n$
- ◊
$$\mathbf{x} = p_1\mathbf{z}_1 + p_2\mathbf{z}_2 + \dots + p_n\mathbf{z}_n \quad (\text{for } p_1, p_2, \dots, p_n \text{ coefficients})$$
- ◊
$$\begin{aligned} (\mathbf{M}\mathbf{M}^T)\mathbf{x} &= (\mathbf{M}\mathbf{M}^T)(p_1\mathbf{z}_1 + p_2\mathbf{z}_2 + \dots + p_n\mathbf{z}_n) \\ &= p_1(\mathbf{M}\mathbf{M}^T)\mathbf{z}_1 + p_2(\mathbf{M}\mathbf{M}^T)\mathbf{z}_2 + \dots + p_n(\mathbf{M}\mathbf{M}^T)\mathbf{z}_n \\ &= p_1c_1\mathbf{z}_1 + p_2c_2\mathbf{z}_2 + \dots + p_nc_n\mathbf{z}_n \end{aligned}$$
- ◊ We will use this equation to analyze multiplication by larger powers of $(\mathbf{M}\mathbf{M}^T)$
- ◊
$$(\mathbf{M}\mathbf{M}^T)^k\mathbf{x} = p_1c_1^k\mathbf{z}_1 + p_2c_2^k\mathbf{z}_2 + \dots + p_nc_n^k\mathbf{z}_n$$

If we consider a vector \mathbf{x} in R^n , and we remember that $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_n$ are a basis for R^n , which means that we can find a linear combination with p_1, p_2, p_n and if we find a linear combination we can obtain the vector \mathbf{x} . This means that every vector can be obtained in terms of these orthogonal eigenvectors and if we expand $\mathbf{M}\mathbf{M}^T$ for such a vector \mathbf{x} we can rewrite the equation and find that $\mathbf{M}\mathbf{M}^T$ values can be substituted by the corresponding eigenvalues c . This equation will be used to analyze multiplication by larger powers of $\mathbf{M}\mathbf{M}^T$

64.5 Convergence of the Hub Authority Computation

vector of hub scores at step k :

1

$$\mathbf{h}^{<k>} = (\mathbf{M}\mathbf{M}^T)^k \cdot \mathbf{h}^{<0>}$$

$$\mathbf{h}^{<0>} = q_1\mathbf{z}_1 + q_2\mathbf{z}_2 + \dots + q_n\mathbf{z}_n$$

$$\mathbf{h}^{<k>} = c_1^k q_1 \mathbf{z}_1 + c_2^k q_2 \mathbf{z}_2 + \dots + c_n^k q_n \mathbf{z}_n$$

Let's divide both sides by c_1^k

2

$$\frac{\mathbf{h}^{<k>}}{c_1^k} = \frac{c_1^k q_1 \mathbf{z}_1}{c_1^k} + \frac{c_2^k q_2 \mathbf{z}_2}{c_1^k} + \dots + \frac{c_n^k q_n \mathbf{z}_n}{c_1^k}$$

$$\text{assumption: } c_1 > c_2 \Rightarrow \lim_{k \rightarrow \infty} \left(\frac{c_2}{c_1} \right)^k = 0$$

$$\lim_{k \rightarrow \infty} \frac{\mathbf{h}^{<k>}}{c_1^k} = q_1 \mathbf{z}_1$$

We reasoned on how to exploit the spectrum of the eigenvectors of the Mmt symmetric matrix, now we apply this to the update rules.

$H^{<k>}$ is the product of $MM^T * h^{<0>}$, we expand $h^{<0>}$ and we find a linear combination of this eigenvector and hence we can do the same for $h^{<k>}$ as we did just before.

When we divide both sides by c_k and implement the assumption of $c_1 > c_2$, which means that the limit to infinity of $(c_2/c_1)^k = 0$ since c_2 is lower than c_1 . If we consider the limit of k that tends to infinity of $h^{<k>} / c_1^k$ is $q_1 \mathbf{z}_1$.

64.6 Final Considerations

(i) a limit in the direction of \mathbf{z}_1 is reached regardless of initial values of $\mathbf{h}^{<0>}:$
let's suppose that $\mathbf{h}^{<0>} = \mathbf{x}$ and that is a positive vector:

$$\mathbf{x} = p_1 \mathbf{z}_1 + p_2 \mathbf{z}_2 + \dots + p_n \mathbf{z}_n, \quad \Rightarrow (\mathbf{M}\mathbf{M}^T)^k \mathbf{x} = c_1^k p_1 \mathbf{z}_1 + c_2^k p_2 \mathbf{z}_2 + \dots + c_n^k p_n \mathbf{z}_n$$

$$\lim_{k \rightarrow \infty} \frac{\mathbf{h}^{<k>}}{c_1^k} = p_1 \mathbf{z}_1$$

(ii) coefficient p_1 (or q_1) must be $\neq 0$: assuring that $p_1 \mathbf{z}_1$ (or $q_1 \mathbf{z}_1$) are non zero vectors, in the direction of $\mathbf{z}_1 \Rightarrow$ textbook

(iii) relax assumption: $c_1 > c_2$

in general we can have $l > 1$ eigenvalues s.t. $c_1 = c_2 = \dots = c_l$, until we find that $c_1 > c_{l+1}$

$$\begin{aligned} \frac{\mathbf{h}^{<k>}}{c_1^k} &= \frac{c_1^k q_1 \mathbf{z}_1 + c_2^k q_2 \mathbf{z}_2 + \dots + c_l^k q_l \mathbf{z}_l}{c_1^k} + \frac{c_{l+1}^k n^k q_{l+1} \mathbf{z}_{l+1} + \dots + c_n^k q_n \mathbf{z}_n}{c_1^k} \\ &\quad k \rightarrow \infty \\ &= q_1 \mathbf{z}_1 + q_2 \mathbf{z}_2 + \dots + q_l \mathbf{z}_l + 0 \end{aligned}$$

this is still a convergence

(iv) authority values: the argument is very similar to hub values (multiplication by $\mathbf{M}^T \mathbf{M}$)

65 PAGE RANK

Page Rank is a generalization of HITS since in HITS we have hub and authority values for each nodes in order to find nodes that are hubs and bring to other nodes with outgoing links to relevant nodes, and other nodes are more authority nodes with a lot of incoming links that could be interpreted as endorsements of the importance of such authorities given the topic we are looking for. If we keep the idea of viewing endorsement as passing directly from one important node to another we can say that endorsements received by in-links and passed across outgoing links.

The definition of page ranking is the following:

- **Step 0:** Init all the pages p to a $PR(p) = 1/n$, where n is the number of pages
- **Step k:** Update all the $PR(p)$ to the sum of all the receiving PR values, normalized by out-links

If we input a query as a keyword we get a lot of pages as a result and these pages could link to each other, so we don't consider the fact that authorities compete with each other, but we just consider the number of incoming and outgoing links. We consider the numbers of ingoing and outgoing links and we get the values of page rank of a certain page

65.1 Page Rank and Stabilization

PR values of all the nodes converge when k tends to infinity but some **degenerate cases** don't. We can reach an equilibrium when if we apply our PR update rule, our limiting values do not change. The problem with degenerate cases is that in some networks some nodes receive all the PR values of the network and hence we apply the PR update rule until we reach an equilibrium. **These degenerate cases are often found in the Out Component of the Web.**

This is a problem since we do not have a path back to some other nodes, and the solution to this problem is to force this stream to flow back to the other nodes sometimes and is done in this steps:

- select a **scaling factor** (aka damping factor) s : s belonging to $[0,1]$, so it is just a probability
- get a portion s of PR values from in-links and then add $(s - 1) / n$
- Now we have convergence for k that tends to infinity
- typically s belongs to $[0.8,0.9]$

66 SPECTRAL ANALYSIS OF PAGE RANK

At step 0 (init):

$$\forall i : r_i = \frac{1}{n}; \text{ n: # pages} \quad r_i = PR(i)$$

At step k:

$$\forall i : r_i = \sum_{j=1}^n M_{ji} \frac{r_j}{k_j^{\text{out}}} \quad (\text{basic PR update rule})$$

$$\forall i : r_i = s \cdot \sum_{j=1}^n M_{ji} \frac{r_j}{k_j^{\text{out}}} + (1 - s) \cdot \frac{1}{n} \quad (\text{scaled PR update rule})$$

At step 0 we initialize our page rank value r_i , that for every i is $1/n$ where n is the number of the pages.

At step k we have that for every i , r_i is equal to the sum of M_{ij} (so if we have a link from j to i) multiplied by the page rank of j (r_j) divided by the out degree of j (k_j^{out}). The idea is that the page rank for j is split up for all the outgoing links of j . If we want to apply the scaled PR update rule we add $(1 - s) * 1/n$

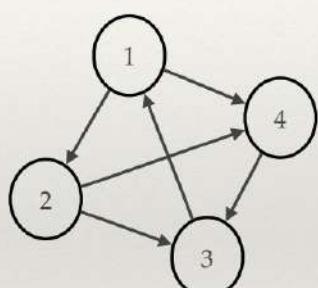
66.1 Using Matrix Notation

\mathbf{N} : Matrix derived from \mathbf{M}

nodes: $1, \dots, n$

$\mathbf{N} : n \times n$

$$N_{ij} = \begin{cases} \frac{1}{k_i^{\text{out}}}, & \text{if } (i, j) \\ 1, & \text{if } (i, j) \text{ is an edge, and } k_i^{\text{out}} = 0 \\ 0, & \text{otherwise} \end{cases}$$



$$\begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

N_{ij} : the share of i 's PR that j should get in one update step

66.2 Update Rule

1 Basic update rule:

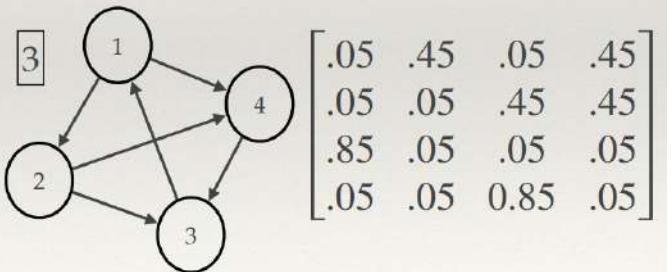
$$\forall i : r_i = \sum_{j=1}^n N_{ji} r_j \\ \leftarrow N_{1i}r_1 + N_{2i}r_2 + \dots + N_{ni}r_n \\ \mathbf{r} \leftarrow \mathbf{N}^T \cdot \mathbf{r}$$

2 Scaled update rule (factor s):

$$\widetilde{N}_{ij} = s \cdot N_{ij} + (1 - s) \cdot \frac{1}{n}$$

4 Application of scaled update rule:

$$\forall i : r_i = \sum_{j=1}^n \widetilde{N}_{ji} r_j \\ \leftarrow \widetilde{N}_{1i}r_1 + \widetilde{N}_{2i}r_2 + \dots + \widetilde{N}_{ni}r_n \\ \mathbf{r} \leftarrow \widetilde{\mathbf{N}}^T \cdot \mathbf{r}$$



66.3 Repeated Improvement

$$\mathbf{r}^{<0>} = \left(\frac{1}{n}, \dots, \frac{1}{n} \right), \text{ initial PR vector}$$

$$\mathbf{r}^{<k>} = (\widetilde{\mathbf{N}}^T)^k \cdot \mathbf{r}^{<0>}$$

Limiting vector $r^{<*>}$ satisfies $\widetilde{\mathbf{N}}^T \cdot \mathbf{r}^{<*>} = 1 \cdot \mathbf{r}^{<*>}$

$\mathbf{r}^{<*>}$ should be an eigenvector of $\widetilde{\mathbf{N}}^T$ with corresponding eigenvalue of 1 BUT $\widetilde{\mathbf{N}}^T$ is not symmetric: this means that eigenvalues can be complex numbers and eigenvectors have no relationships to one another

66.4 Convergence of the Scaled PR Update Rule

$$\forall i, j : \widetilde{N}_{ij} > 0$$

Perron's theorem

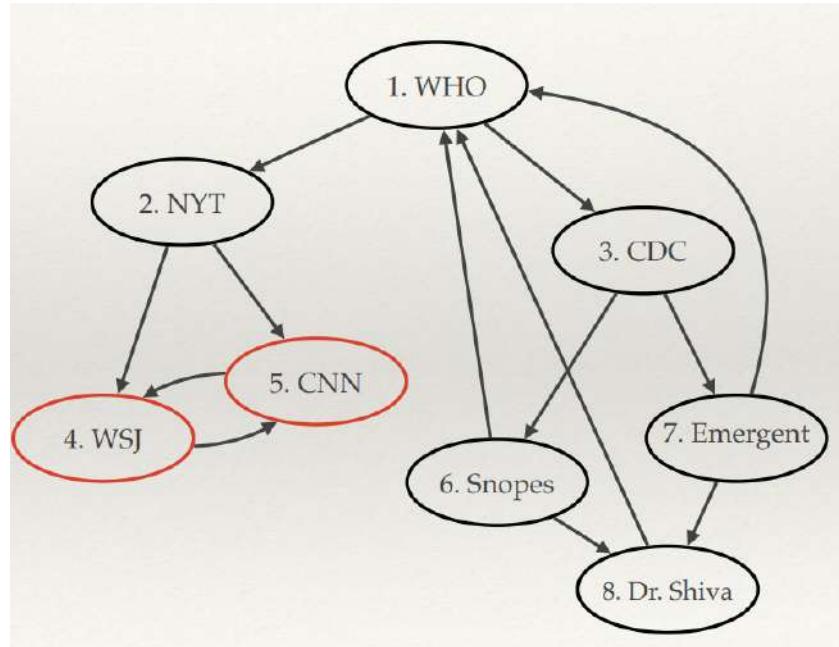
Matrix \mathbf{P} (with entries > 0)

- i) \mathbf{P} has an eigenvalue $c > 0$ s.t. $c > c'$ $\forall c'$ (with c' another eigenvalue)
- ii) Exists an eigenvector \mathbf{y} with real positive values corresponding to c , and \mathbf{y} is unique (up to a multiplication constant)
- iii) if $c = 1$, then for any starting vector $\mathbf{x} \neq \mathbf{0}$ with non negative coordinates, the sequence of vectors $p^k \mathbf{x}$ converges to a vector in the direction of \mathbf{y} ($k \rightarrow \infty$)

67 RANDOM WALKS AND PAGE RANK

Random Walks is just randomly clicking from one page to another, picking each page with equal probability. The idea is to follow links for a sequence of length k , just as a drunk wandering dude would do in a city. We need to make a claim though, that the probability of being at page x after k steps is the application of the basic PR update rule. And a more in depth intuition allows us to say that $PR(x)$ is the limiting probability that a random walk across hyperlinks will end up at x as we sum the walk for larger and larger number of steps.

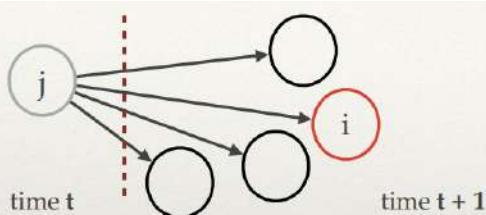
67.1 Leakage



If we are in a situation where we can only go back and forth between two nodes, we are stuck forever.

The solution is that with a probability s , the random walker clicks on an hyperlink in the page and with probability $1 - s$ he goes to a randomly selected node in the network, such as clicking a random site in my bookmark bar or inputting a site in the site bar

67.2 Formulation of The Page Rank using Random Walks



Which is the probability of being at node i at time $t+1$?

b_1, b_2, \dots, b_n : the probabilities of being at node i in a given step.

$b_i \leftarrow \sum_{j=1}^n M_{ji} \frac{b_j}{k_j^{out}}$: the probability of being at node j in the following step

Using matrix \mathbf{N} : $b_i \leftarrow N_{1i}b_1 + N_{2i}b_2 + \dots + N_{ni}b_n \Rightarrow \mathbf{b} \leftarrow \mathbf{N}^T \cdot \mathbf{b}$

claim: PR of page i is exactly the prob of being at node i after k step. (qed) \square

67.3 A scaled Version of the Random Walk

For a given probability s : the walker follows a random outgoing edge

With prob $(1-s)$: the walker is teleported uniformly at random to another node

$$b_i \leftarrow s \cdot \sum_{j=1}^n M_{ji} \frac{b_j}{k_j^{out}} + \frac{(1-s)}{n}$$

Using matrix \widetilde{N} : $b_i \leftarrow \widetilde{N}_{1i}b_1 + \widetilde{N}_{2i}b_2 + \dots + \widetilde{N}_{1n}b_n \Rightarrow \mathbf{b} \leftarrow \widetilde{\mathbf{N}}^T \cdot \mathbf{b}$

claim: PR is equivalent to the scaled version of random walks. (qed) \square

68 INFORMATION CASCADES - FOLLOWING THE CROWD

68.1 The Milgram Experiment on Obedience

Milgram, who we have met on the six degree of separation notion, made an experiment where some subjects were taking orders from an authority to press a button that sends an electrical shock to a person that has committed a crime. Very few elements were given to the subject regarding how the felon had committed the crime, or even if he had committed it at all. The point is that the more the subject pressed the button, the higher the electrical shock sent to the felon would be, but of course the felon was just an actor.

The key point is that the maximum intensity of the electrical shock could have caused death to the felon, and after a few sessions of the authorities given orders to the subjects, an high percentage of the subjects where pressing the button knowing that they would send out an electrical shock that would kill the felon, and this was a pretty impressive result, since the phrase 'we were just following orders' could be interpreted in many bad ways.

These experiments let us think that people are directly influenced by others and induced behaviors can be against common sense or even against our own moral ideas.

But can we even be influenced by other factors that are apparently irrational?

68.2 Collective Behavior

Robert Park coined **Collective Behavior** because it's a matter of fact that not only an individual can be pushed by another individual to do something, but individuals can be influenced by a collective and more general behavior. Collective behavior is defined as "the behavior of individuals under the influence of an impulse that is common and collective, an impulse, in other words, that is the result of social interaction".

So not only a person with law authority can influence some one, but even groups of people with strong ideas such as cults, crowds etc.. can influence us into doing something that we wouldn't normally do. So collective behaviors are the activities of people in large groups such as cults or publics.

68.3 The Milgram Experiment on Conformity and Independence

This is an experiment that proves how easy we could set up a social experiment in our life. If we go outside and just start stargazing at the sky in the middle of a crowded square, after a while a lot of people will start and look up at the sky just to see what is going on. It is enough to have one or two people looking up to make a whole crowd start looking up as well.

A classic interpretation for this is that there is a **social force for conformity** that grows stronger as the group conforming to the activity becomes larger. But if we give an interpretation based on information cascades we can say that with more and more people looking up, future passerby may have rationally decided that there was good reason to also look up and that is because perhaps those looking up have access to some information that we do not have, even if this may contradict what we possibly already know, we can decide to change behavior, and this is not quite as irrational as we might think.

68.4 Information Cascade

Herd or information cascade occurs when people abandon their own information in favor of inferences based on earlier people's actions. This is different from situations where we can have a direct-benefit effect, since we align our behavior to others' behavior because we may have a direct benefit that we feel like we need. An example could be buying the same phone or clothes as my friends or colleagues.

69 A SIMPLE HERDING EXPERIMENT

69.1 Asch Conformity Experiment

We start from this to get an understanding, in this experiment a group of people is asked which is the shortest of three lines and a group of people claimed that the second shortest line was the shortest, but when a real group of subjects comes into the experiment and obviously sees that the answer is not the previous, still tends to give it as an answer since there is a tendency to conform to the most common answer, even if wrong.

69.2 Herding Experiment (Herding, urns and marbles)

We want to understand with a simple experiment if a simple rational behavior can explain this apparent lack of rationality while making a decision, the ingredients of this experiment are:

- A decision to be made
- People make the decision sequentially, and choice is public
- Each person has some private information that helps guide their decision
- A person cannot observe the others' private information directly, but can try to infer this information from their decision

We have an urn and we have two alternate universes and we don't know before hand which is the universe that we will be living in. In the urn that is inside of one universe we have two red tokens and one blue, while on the other universe we have two blue tokens and one red token. Of course the probability to choose the universe with the 2 red tokens is 50 percent, and the same for the one with the two blue tokens, but no one knows in which of these two universes we are living.

Each participant can draw one token (marble), look at its color privately, put the marble in the urn again, then make a guess between the two options, and announce the guess to the others. To make the argument stronger we let's assume that every participant is honest, and that they will make a guess based on rational basis and won't lie about their guess and don't want to deceive the others.

If the first person sees that the marble is blue, it is better to guess that the urn is majority-blue and hence the first person says aloud to the audience that he thinks that the universe he is in is the majority blue. Even if this person didn't say aloud which was the color of the token he got, he makes his guess available to every one saying that the urn in this universe is a majority blue. This is the same as saying that he took from the urn a blue token, without revealing that he actually got a blue token, but just making it known from his decision of saying that the urn is a majority blue.

If also the second person takes a blue token from the urn it is natural to guess that the urn is majority blue.

The third person knows that the first two guesses have perfect information since the two first persons were honest about which color they picked from the urn. It is like drawing three times from the urn. If the two guesses are the same, regardless of which color the player sees, the best strategy would be to bet on majority blue. So, no matter which is the color of the token of the third player, because even if the token is red, he knows that the two previous persons have said that the urn is majority blue, and hence he says that the urn is majority blue.

From the fourth person onward the rational reasoning is the same as the third player since they know that the first two guesses convey perfect information. They also know that from player 3 on, regardless of their own private information, they would say that the urn was majority blue, so their guesses convey no usable information and **an information cascade starts** since no one is under the illusion that every single person is drawing a blue marble, but once the first two guesses turn out blue, the future announced guesses become useless and so everyone's best strategy is to rely on the limited perfect information they have available from the first drawings of the tokens.

If we open up the urn and we have a majority red, we would draw a blue token with a probability of $1/3$. The second player, since the token is put back into the urn, would still have the same $1/3$ probability to draw a blue token, and hence the probability that two players would draw the blue token is $1/9$. In this scenario an information cascades can lead to nonoptimal outcomes and there is only a $1/9$ chance that an information cascade would start with people betting on the wrong majority.

We could introduce the concept of **wisdom of the crowd**, what if we have more players? **Nothing will change**, because everyone will guess blue if just the first two players will guess blue.

If we consider another eventuality where the first player draws a blue token, the second and the third draw a red token, the fourth and fifth a blue token etc.. we need a more complex and sophisticated model to predict how many similar tokens we need to draw in a sequence from the urn in order to set the information cascade in motion to start.

Information cascades are quite easy to be interrupted and hence are very fragile, since a simple change in the pattern of drawn tokens could break the cascade. We could be in a situation where the fist two player play majority blue, and hence every one after plays majority blue for a cascade effect, but if at a certain point two players draw a red token and show it to every other player, then the next player would make his choice according to four pieces of perfect information (the blues from the beginning and the reds from the draws before him). The player could then bet on the majority red with quite come certainty, even if all the players before him have bet on a majority blue urn by just following the crowd.

70 BAYES RULE – A model of Decision Making under Uncertainty

$P[A | B]$ conditional probability of A given B

$$P[A | B] = \frac{P[A] \cdot P[B | A]}{P[B]}$$

$P[A]$ prior probability of A

$P[A | B]$ posterior probability of A given B

- you have been **tested positive** for a terrible and rare disease (event: E)
- 0.1% of people **has this terrible disease** (event: H , $P[H]$)
- 99% of people that have the disease will be **tested positive** ($P[E|H]$)
- Which is **the chance that you got this disease** ($P[H|E]$)?

$$P[H|E] = \frac{P[H] \cdot P[E|H]}{P[E]} = \frac{P[H] \cdot P[E|H]}{P[H] \cdot P[E|H] + P[-H] \cdot P[E|-H]}$$

$$= \frac{.001 \cdot .99}{.001 \cdot .99 + .999 \cdot .01} = 9\%$$

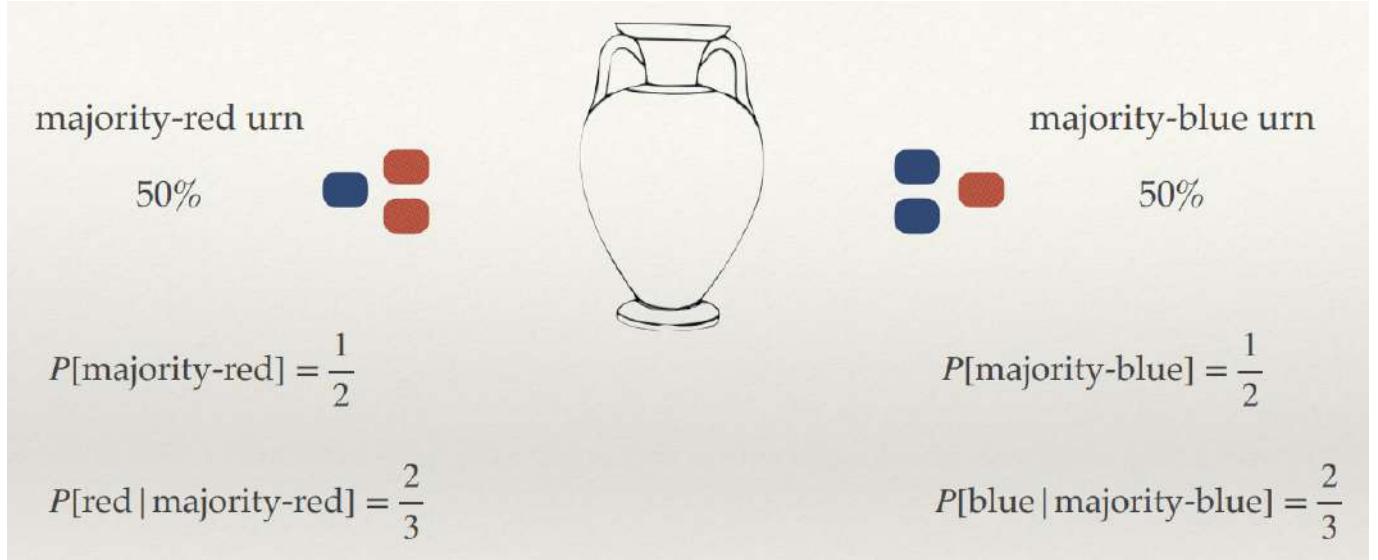
- Now, you want a second independent opinion
- Numbers are the same, but for the *prior probability* of your hypothesis $P[H] = 9\%$

$$P[H|E] = \frac{P[H] \cdot P[E|H]}{P[E]} = \frac{P[H] \cdot P[E|H]}{P[H] \cdot P[E|H] + P[-H] \cdot P[E|-H]}$$

$$= \frac{.09 \cdot .99}{.09 \cdot .99 + .91 \cdot .01} = 91\%$$

70.1 Bayes Rule in the Herding Experiment

If we apply the Bayes Rule in the Herding Experiment we can define the probabilities for the majority red and blue urn:



If the first player extracts a blue token, we can try and compute the probability of having a majority blue urn given that we extracted a blue token. The difficult part is to compute the probability of drawing a blue token, and we can compute it as the probability of the majority blue multiplied by the probability of drawing a blue given that we have a majority blue urn, plus the probability of being in a majority red universe multiplied by the probability of drawing a blue token given that we are in a majority red universe

$$P[\text{majority-blue} | \text{blue}] = \frac{P[\text{majority-blue}] \cdot P[\text{blue} | \text{majority-blue}]}{P[\text{blue}]} \quad \text{player n}$$

$$= \frac{\frac{1}{2} \cdot \frac{2}{3}}{P[\text{blue}]} = \frac{\frac{1}{3}}{P[\text{majority-blue}] \cdot P[\text{blue} | \text{majority-blue}] + P[\text{majority-red}] \cdot P[\text{blue} | \text{majority-red}]}$$

$$= \frac{\frac{1}{3}}{\frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3}} = \frac{2}{3} \quad \Rightarrow \text{Guess: majority-blue}$$

The final probability of being in a majority blue universe given that we drew a blue token is $2/3$, and the rational guess for player 1 is to say majority blue. We can then apply the same reasoning to the second player if he draws the blue token, and hence we have a majority blue.

If the third player draws a red token, we want to compute the probability of being in a majority blue universe knowing that we drew a blue, blue and red token.

$$\begin{aligned}
 P[\text{majority-blue} | \text{blue, blue, red}] &= \frac{P[\text{majority-blue}] \cdot P[\text{blue, blue, red} | \text{majority-blue}]}{P[\text{blue, blue, red}]} = \frac{P[\text{majority-blue}] \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3}}{P[\text{blue, blue, red}]} = \frac{\frac{1}{2} \cdot \frac{4}{27}}{P[\text{blue, blue, red}]} \\
 &= \frac{\frac{2}{27}}{\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}} = \frac{2}{27} \cdot 9 = \frac{2}{3} \quad \Rightarrow \text{Guess: majority-blue}
 \end{aligned}$$

Using again the Bayes rule we compute this probability by multiplying the probability of having a majority blue urn with the probability of extracting a blue,blue,red combination of tokens knowing that we have a majority blue urn and divide it by the probability of extracting blue,blue and red token in sequence. For the denominator we use the same idea of before, so we multiply the probability of being in a majority blue universe ($1/2$) for the probability of having the blue, blue and red sequence given the majority blue, and then we sum it to the the probability of being in a majority red universe ($1/2$) multiplied by the probability of having a blue, blue, red sequence given that we are in a majority red universe. We get a $2/3$ probability and still the best bet for the third player could be to bet on majority blue, and so the third player won't have a good chance on betting on majority red independently on the private information that he has regarding his drawing of a red token after only two previous draws.

71 A SIMPLE CASCADING MODEL

71.1 Model Goals

We want to model situations where people make decisions randomly and sequentially. Decisions made on a combination of private information and observations of what earlier people have done are the basic ingredients of a model. **We will use the Bayes Rule to predict that cascades will form with probability tending to 1 as the number of people goes to infinity**

71.2 Model Ingredients and Assumptions

- ◊ Group of **people**: $1, 2, 3, \dots, n$
 - ◊ they sequentially make decisions: they have to **accept** (A) or **reject** (R) a given *option*
 - ◊ ex., A if they guess that the urn is "*majority-blue*", R otherwise
 - ◊ **States** of the world: the given option can be **good** (G) with probability p or **bad** (B), with probability $1-p$
 - ◊ ex., "*majority-blue*" can be our G state of the world, randomly assigned with prob. $\frac{1}{2}$
 - ◊ **Payoffs**: each individual receives a payoff after their decision.
- $$\text{payoff} = \begin{cases} 0 & \text{if guess is } R \\ v_g > 0 & \text{if guess is } A \wedge \text{ state is } G \\ v_b < 0 & \text{if guess is } A \wedge \text{ state is } B \end{cases}$$

- ◊ We also assume that the expected payoff from accepting in the absence of other information is equal to 0: $v_g p + v_b (1-p) = 0$
 - ◊ i.e., the expected payoff from accepting is the same as the payoff from rejecting, before an individual gets any additional information
- ◊ **Signals**: before guessing, every individual gets a private signal. The signal can be **high** (H) - suggesting acceptance - or **low** (L) - suggesting rejection
 - ◊ if accepting is a good idea, then high signals are more frequent, otherwise low signals are more frequent

- ◊ this can be formalized by means of a frequency $q > \frac{1}{2}$

- ◊ Ex., If a player draws a "*blue*" marble from the urn, and the urn is actually "*majority-blue*", then it is a *high signal* with frequency $\frac{2}{3}$

		States	
		B	G
Signals	L	q	1-q
	H	1-q	q

71.3 Final Considerations of the Herding Experiment

- ❖ We have two states of the world: "majority-blue" (G) and "majority-red" (B), each created with probability:
these are prior probabilities $P[G] = p = \frac{1}{2} = 1 - p = P[B]$
- ❖ When the player draws a marble, and it is blue, then the signal is H:
We have posterior probabilities $P[H|G] = q = \frac{2}{3}, P[H|B] = 1 - q = \frac{1}{3}$
- ❖ We can reformulate our experiment to find, for each individual, the outcome of $P[G|H]$ to make a decision between accepting or rejecting option G

71.4 Individual Decisions

People generally make decisions about accepting or rejecting and we can model this on the basis of private information (signals) only. We can then compute the probability of being in a good state given that we have an high signal.

- ❖ Now we can model how people make decisions about accepting or rejecting
- ❖ Let's do that on the basis on private information (signals) only

$$\begin{aligned} P[G|H] &= \frac{P[G] \cdot P[H|G]}{P[H]} \\ &= \frac{P[G] \cdot P[H|G]}{P[G] \cdot P[H|G] + P[B] \cdot P[H|B]} \\ &= \frac{p \cdot q}{p \cdot q + (1-p) \cdot (1-q)} \end{aligned}$$

$$(*) \quad > p$$

In fact, a high signal is more likely to occur if the option is good!

An individual that observes an high signal should raise their estimate

We want now to prove that this is greater than p

$$\begin{aligned} &= p \cdot q + (1 - p) \cdot (1 - q) \\ &< p \cdot q + (1 - p) \cdot q \\ &= q \\ &\Rightarrow \frac{p \cdot q}{p \cdot q + (1 - p) \cdot (1 - q)} \\ &> \frac{p \cdot q}{q} \\ &= p \end{aligned}$$

This means that we have that the probability of having a good state given an high signal is greater than the probability a priori, in other words a posteriori probability is higher than a priori probability.

71.5 What about the Payoff?

- ❖ Before signal and without other information: payoff = 0
 - ❖ $v_g P[G] + v_b P[B] = 0$
- ❖ Now, we can raise it to:
 - ❖ $v_g P[G|H] + v_b P[B|H] > 0$
- ❖ They should **accept** the option
- ❖ Analogous reasoning if the player observes a low signal: it will be better rejecting it

71.6 Multiple Signals

We know that the more information we get, the more predictive Bayes' Rule gets. An individual takes a decision when they are exposed to a sequence S of independently generated signals consisting of a high signals and b low signals. In our urn example we have that when player 3 observes the sequence $S=[\text{blue}, \text{blue}, \text{red}]$ and the option to be accepted or rejected is majority-blue, then we have $a = 2$, and $b = 1$

$$\begin{aligned}
 P[G|S] &= \frac{P[G] \cdot P[S|G]}{P[S]} && \text{signals are generated independently, hence} \\
 &= \frac{p \cdot q^a(1-q)^b}{P[S]} && \text{we can multiply their probabilities} \\
 &= \frac{p \cdot q^a(1-q)^b}{P[G] \cdot P[S|G] + P[B] \cdot P[S|B]} && \text{we have to expand out the two possible} \\
 &&& \text{ways of getting that sequence} \\
 &= \frac{p \cdot q^a(1-q)^b}{p \cdot q^a(1-q)^b + (1-p) \cdot (1-q)^a q^b} && \text{signals are independent, again, and we have} \\
 &&& \text{to factor the probability of } P[H|G] \text{ } a \text{ times,} \\
 &&& \text{and the probability of } P[L|G] \text{ } b \text{ times} \\
 &&& \left\{ \begin{array}{l} < p \\ = p \\ > p \end{array} \right. ?
 \end{aligned}$$

First, focus on the denominator: $p \cdot q^a(1-q)^b + (1-p) \cdot (1-q)^a q^b$

$$\begin{aligned}
 \text{If we substitute it with:} \quad & p \cdot q^a(1-q)^b + (1-p) \cdot q^a(1-q)^b \\
 &= q^a(1-q)^b \\
 \Rightarrow & \frac{p \cdot q^a(1-q)^b}{q^a(1-q)^b} \\
 &= p
 \end{aligned}$$

denominator	substitution
$p \cdot q^a(1 - q)^b + (1 - p) \cdot (1 - q)^a q^b$	$p \cdot q^a(1 - q)^b + (1 - p) \cdot q^a(1 - q)^b$

(i) if $a > b$, then makes the denominator smaller (remember that $q > \frac{1}{2}$). Then we have that expression is larger than p . $P[G | S] > p$
accept G

(ii) if $a < b$, the argument is symmetric: $P[G | S] < p$
reject G

(iii) if $a = b$, the denominator and the substitution are the same: $P[G | S] = p$
indifferent to G or B

72 SEQUENTIAL DECISION MAKING AND CASCADES

We know what individuals decide, but not what they really know. Players have access to their own private signal and also to accept or reject decisions of the earlier people.

The first person just follows his own private signal, that is an high signal. We can draw in a cartesian axis system on the y axis the subtractions of low signals to high signals ($a-b$) according to the sequence of people that observe their own signals.

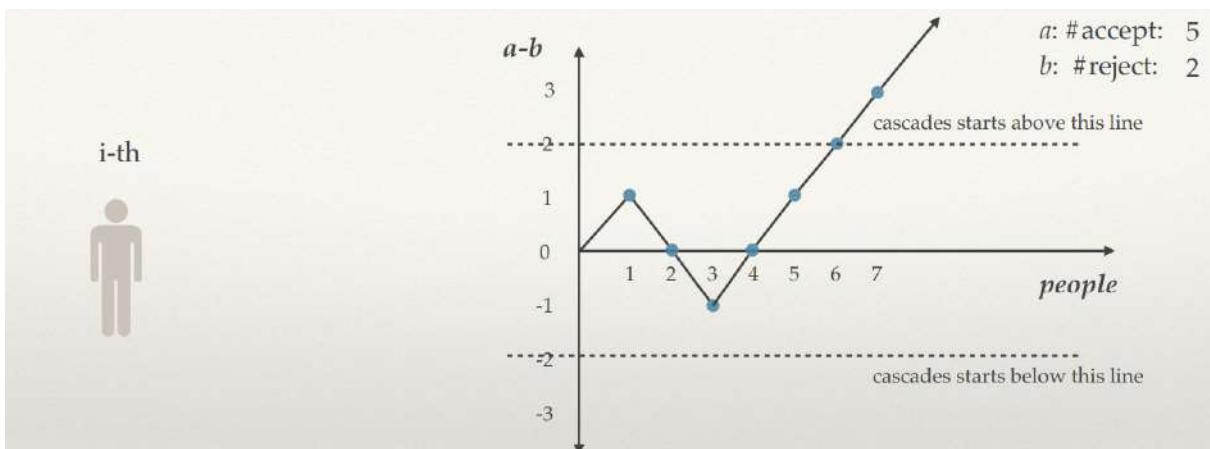
Also the second person just follows their private signal, even if it does not match what observed so far. In other words we assume that if indifferent, then the player decides according their own signal.

The third person has a sequence of three independent signals, which the two previous signals are contradictory, but $a = b$, and than the player is deciding according their own information again

The fourth person will find $b > a$, and their own signal will make them indifferent again. The player accepts following their own signal

The fifth and sixth person both get an high signal and follow their decision, but when the seventh player has to make his decision something changes

When the seventh player receives a low signal he needs to make a decision but $a > b$ and the player signal should suggest rejection. However, actual private signal cannot change this earlier majority and it is not enough for making the decision over acceptance indifferent again. This player and the ones after him will follow the crowd regardless their own signal and hence a cascade starts.



Observe that information cascades start when $|a - b| \geq 2$.

This corresponds of a sequence of 3 identical signals

This proves that under rational behavior, information cascades starts when the number of high signals subtracted with the low signals is equal or greater then 2. And this corresponds to a sequence of three identical signals.

72.1 What Happens when the Number of Individuals Grows to Infinity?

Let's prove that the probability of finding three matching signals in a row converges to 1

Players: $1, 2, 3, 4, 5, 6, \dots, n$ Group them in sequences of three: $(1, 2, 3), (4, 5, 6), (7, 8, 9), \dots$

The probability that *one of this group* has three matching signal is: $q^3 + (1 - q)^3$

The probability that *none of these blocks* consists of identical signals is therefore:
 $(1 - q^3) - (1 - q)^3)^{n/3}$

$$\lim_{n \rightarrow \infty} (1 - q^3) - (1 - q)^3)^{n/3} = 0$$

We are able to predict that **cascades will form** with probability tending to 1 as the number of people goes to infinity.

73 LESSONS FROM CASCADE EFFECTS

- We are able to predict that cascades will form with probability tending to 1 as the number of people goes to infinity, and hence we know that even with the most unlucky sequence of numbers, sooner or later we will witness the formation of a cascade.
- **Cascades can be wrong** because we could be just following a crowd and there is no guarantee that my rational behavior, in terms of probability, could make a more precise guess.
- **Cascades can be based on very little information**, even if we have an infinite amount of signals we could make a mistake since cascades are based on a very short sequence of signals that start from little amounts of information
- **Cascades are quite fragile** because some participants of the cascade can make public their own private informations and values and then a new cascade could form

If we want to really exploit the **Wisdom of the crowds it is necessary to prevent information cascades from forming and hence participants must guess independently**, without knowing what the others have guessed because the possibility of other people to influence other is quite strong.

74 NETWORK EFFECTS

There are different reasons why individuals may imitate others:

- **Informational effects**, also against the evidence of one's own private information
- **Network effects**, based on direct-benefit of individuals

We can also have some natural settings such as the **adoption of technologies** for which interaction or compatibility with others is important. If we use a certain device like a phone, the value of these certain devices comes from the number of the people that use this device.

74.1 Externalities

Externality is a situation in which the welfare of an individual is affected by the actions of other individuals, without a mutually agreed-upon compensation. We can have two different types of externalities:

- **negative externalities**: where they cause a decrease in welfare, as in **traffic congestion**
- **positive externalities**: where individuals' welfare increases

positive externalities can have consequences due to network effects, this means that the payoffs depend on the number of others who use a certain device or good and not on the details on how they are connected.

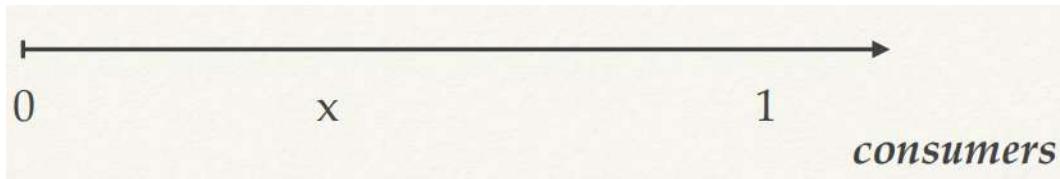
75 THE ECONOMY WITHOUT NETWORK EFFECTS

75.1 The Market for a Good

We will first consider the market functions with no network effect, where consumers do not care how many users of the good there are. So this allows us to focus on markets with a huge number of potential purchasers, each of whom is small enough relative to the entire market that he or she can make individual decisions without affecting the aggregate behavior.

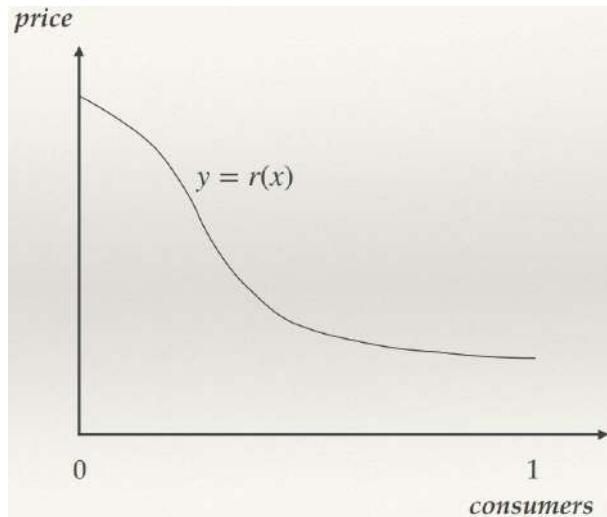
We model this market as the set of real numbers in the interval $]0,1[$:

- c is a consumer that belongs to this interval and hence the total mass of the consumers is 1
- the set of consumers between 0 and $x < 1$ represents an fraction of the population



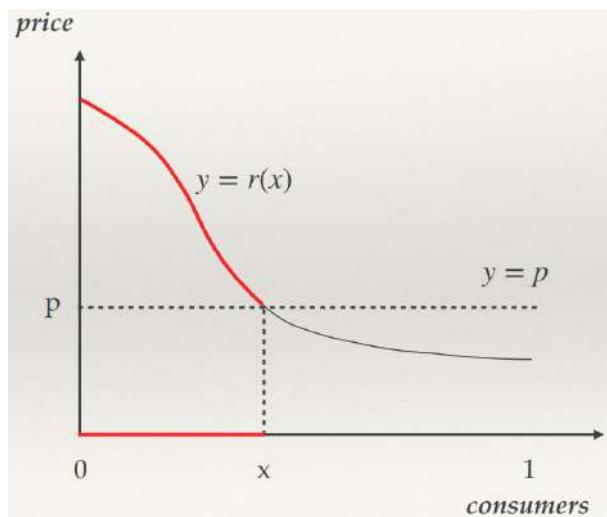
Each consumer wants at most one unit of the good, and an intrinsic interest in obtaining the good, that can vary from one consumer to another. In this model we have **no network effect**, where the consumer's willingness to pay is entirely determined by this intrinsic interest.

75.2 Reservation Prices



Reservation Price $r(x)$ is something that we can put on the y axis and is the maximum amount a consumer x is willing to pay for one unit of the good. We can sort all the consumers such that this reservation price is lower and lower while we approach the total mass of the consumers (1) and hence we have that if $r(x) > r(x')$ then $x < x'$, so basically we have a decreasing reservation price and we want to sort the consumers in order to have a decreasing function.

75.3 Market Price



Reserveation price is the maximum value a person would pay for a good, while market price p is the price for a certain good. No units are offered for sale at a price above or below p , ideally of course.

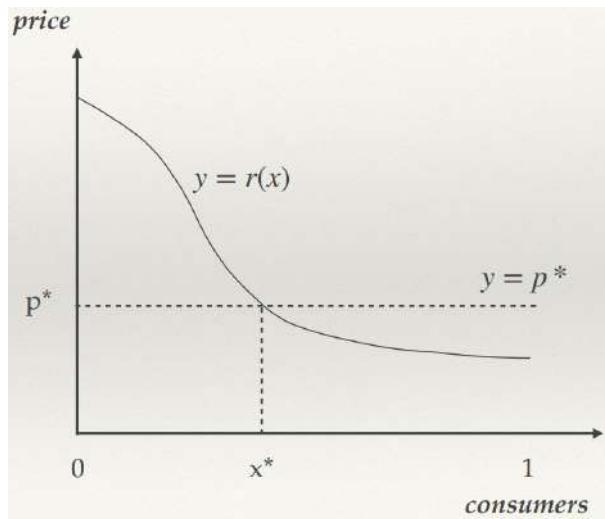
In this generalization we have that there is some unique number x of consumers with the property that the reservation price is equal to the market price ($r(x) = p$).

Hence all the consumers between 0 and x will buy the product because they are giving to the manufacturer more or equal value to the price of the product itself. While no one with a lower value of the product is willing to buy it. Hence the consumers that want to pay less then their reservation price won't buy the product.

The relation between price and quantity is the market **demand** for the good.

75.4 The Equilibrium Quantity of the Good

A certain good can be produced at a constant cost p^* per unit, ideally of course. As for consumers, there are many potential producers of the good so that none of them is large enough to be able to influence the market price of the good. The producers will supply any amount of the good at a price of p^* per unit, and none of the good at any price below p^* . And moreover, the price cannot remain above p^* , since any profit to a producer would be driven to zero by competition from other producers. Since if we decide to supply the amount of the good at a higher price, we won't sell this product to anyone since all the other competitors are selling the same product at a lower price.



We can assume a market price of a **constant cost per unit** p^* , regardless of the number of units of the good produced. And hence we have that the price p^* is included in the minimum reservation cost and the maximum reservation cost, $r(1) < p^* < r(0)$.

The **supply** of the good is x^* such that $r(x^*) = p^*$.

x^* is the fraction of people in the market that will buy the product since their reservation price is lower or equal to product price. Hence x^* is **the equilibrium quantity of the good**, given the reservation prices and the cost p^* .

If less than an x^* fraction of the population purchased the good (say x'), there would be consumers who have not purchased, but who would have an incentive to do so, because of the reservation price above p^* . Now there are a lot of people whose reservation price is higher than the actual price p^* and hence they have an incentive to buy the product, and we have an **upward pressure** since $r(x) > p^*$, and this upward pressure means that the x' consumers will have a tendency to move to the x^* consumers.

If more than an x^* fraction of the population purchased the good (say x''), there would be consumers who have purchased, but who would have an incentive not to do so, because of the reservation price below p^* , and hence we have a **downward pressure** $r(x) < p^*$.

This equilibrium is also socially optimal, since if we consider the **social welfare** of the allocation, which is the sum over x that goes from 0 to x' of the reservation price minus the market price ($r(x) - p$), if consumers between 0 and x' are the consumers that receive

the product. The best possible value for this sum is when x' is when $r(x^*) = p^*$, and in this situation we would have the optimal social welfare and every body is satisfied.

76 THE ECONOMY WITH NETWORK EFFECTS

With network effects, a potential consumer takes into account also the total number of users of the good and hence we need to formulate a **new reservation price** $r(x)f(z)$, where:

- $r(x)$ is the **intrinsic interest** of x in the good, maybe because we like a certain device or product
- $f(z)$ measures the **benefit to each consumer from having a z fraction of the population using the good**, like the other people that use this device or product

$f(z)$ is increasing in z and it dictates and controls **how much valuable a product is when more people are using it**, and hence the multiplication $r(x)f(z)$ means that those who give a greater intrinsic value on the good, will benefit more from an increase of z than those who place a smaller intrinsic value on it.

To make a more precise formalization we need to make some assumptions:

- $f(0) = 0$, which means that if there is no one using the product, $f(z)$ is zero and hence it has no benefit
- $f(1) = 0$, which means that the total mass potential consumer is reached, $f(1)$ is zero and hence if every one has that product then it wont reach anymore consumers in the market

A consumer's willingness to pay depends on the fraction of the population using the good each which means that **each consumer needs to predict what this fraction will be in order to evaluate to purchase or not**. If p^* is the price of the good, and that consumer x expects a z fraction of the population will use the good, then x will want to purchase provided that $r(x)f(z) \geq p^*$. If the reservation cost, that now is the multiplication of two functions, is greater or equal then p^* . we will buy the product.

76.1 Perfect Predictions Hypothesis

Perfect Predictions Hypothesis are a set of hypothesis that are **self-fulfilling expectations equilibrium for the quantity of purchasers z** . It means that the consumers form a shared expectations that the fraction of the population using the product is z , and if each of them then makes a purchasing decision based on this expectation, then the fraction of people who actually purchase is in fact z . If we convince a population that a million persons would buy the product and we still have a 100 thousand people that buy the product, there are a lot of other consumers that will think that the maximum threshold has not been reached because they are pointing to a million purchases and then a lot of people will buy the product since the threshold hasn't been reached yet.

The equilibrium of z , in terms of the price $p^* > 0$ looks like:

- if $z = 0$ the $r(x)f(z)$ will be zero and hence no one will want to purchase, and the shared expectation of will be fulfilled
- if $0 < z < 1$ we have that the set of consumers between 0 and z will purchase the product, and the lowest reservation price in this set will belong to consumer z

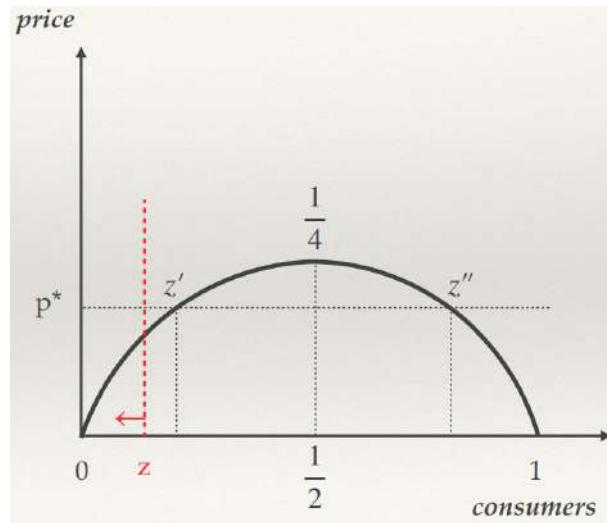
If the price $p^* > 0$ together with the quantity $0 < z < 1$, they form a **self-fulfilling expectation equilibrium**, then $p^* = r(z)f(z)$, since the fraction of the population that bought the

product at equilibrium is z , since $x = z$.

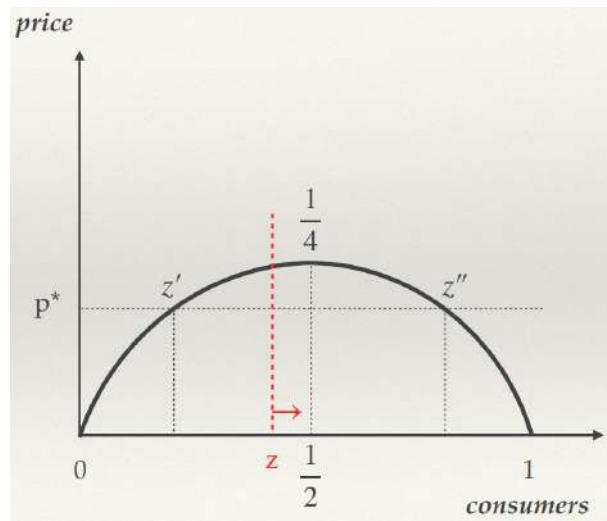
Observe this is a clear contrast with the model with no network effect, where $p^* = r(x^*)$, with r decreasing in x , **hence at high prices the number of units of the good that can be sold is smaller. Markets with network effects are more complicated.**

77 STABILITY, INSTABILITY AND TIPPING POINTS

We can have different equilibriums in our model, each equilibrium with its own properties. In certain areas of the price consumer graph we have equilibriums where people will buy the product since it is equal to their reservation price, while others that won't since the price is higher than their reservation price.

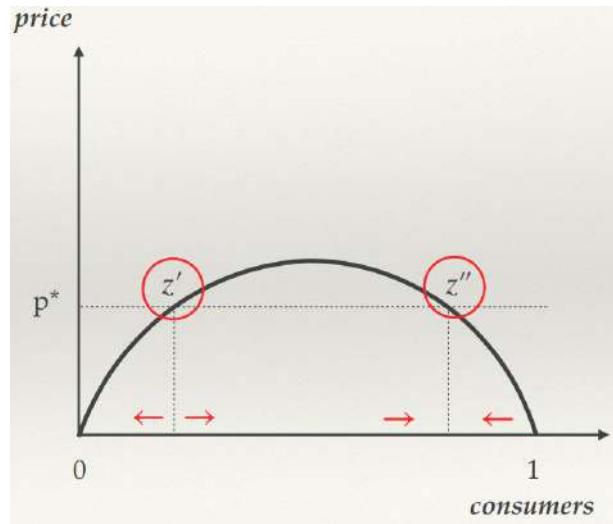


If we suppose that a fraction z (not equilibria) of the population is to purchase the product, and if we have that this fraction is between 0 and z' we can expect a **downward pressure** since the price is higher than my reservation price.



If we suppose instead that this fraction of people z is between z' and z'' we can expect an **upward pressure since** the price we are willing to pay is higher than the actual market price of the product

While if z is greater than z'' we still have a downward pressure since people would want to pay less than the upper cost.



We have strong consequences for non zero equilibrium since if we are on z' we would not move, but if we are a little further on the left from z' we will move out from its radius, while for z'' we have the opposite result since z'' tends to attract clients towards its radius instead of letting them get farther away.

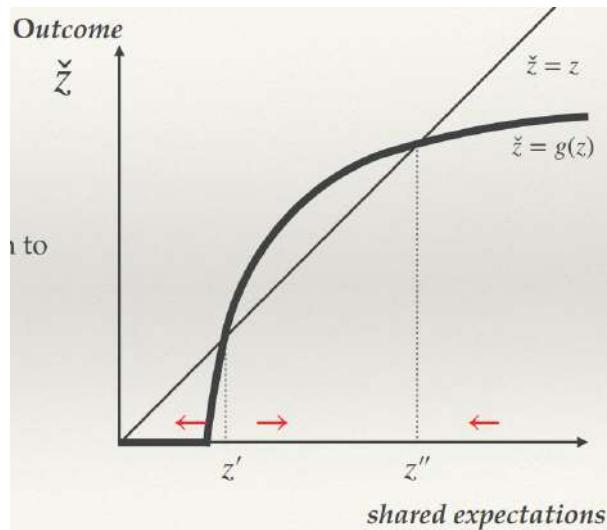
Hence we have that z' is more **unstable** than z'' and it is called a **tipping point**.

Lowering the price p^* has two beneficial effects:

- the low equilibrium would move left, with a lower price and hence easier to purchase for a customer
- the right equilibrium would move right and hence the potential user population gets larger

The relevant thing is that the tipping point z' will be much closer to 0 so the effort that would be needed to lose the balance of this point is lower.

78 A DYNAMIC VIEW OF THE MARKET



With network effects in general (and relaxing the perfect hypothesis) we usually expect to find a curve that looks like an upstanding worm. When the curve $z \text{ cappello} = g(z)$ crosses the line $z \text{ cappello} = z$ we have equilibria that can be stable or unstable depending on whether the curve crosses from above or below the line, hence downward pressure or upward pressure make this point stable.

78.1 Modeling the Population Reaction

Granovetter modeled how a population may react to a network effect by focusing on **dynamic behavior**, that is how the number of people participating in a given activity with network effects would tend to **grow or shrink over time**. An example could be the explosion of a new social network like tiktok etc.. and here we want to formulate this in terms of participating and subscribing to a certain social instead of buying a good.

The model remains pretty much the same as the economical one:

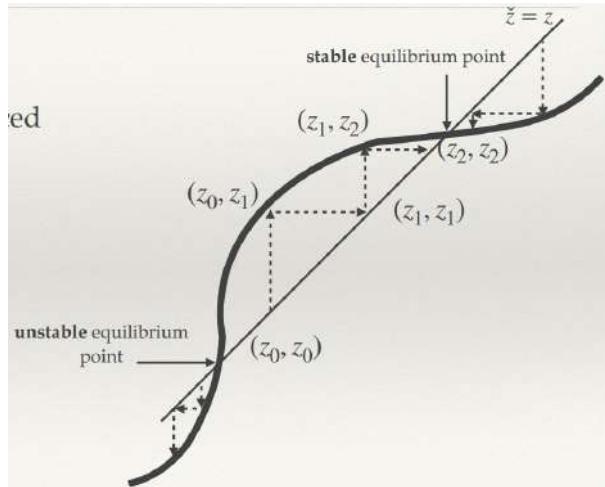
- a person x has an intrinsic interest in using the site ($r(x)$)
- the site is more attractive if it has more users ($f(z)$)
- there is a fixed level of effort required to use the site, which is the price p^*
- if person x expects a z fraction of the population to want to participate, than x will participate if $r(x)f(z) \geq p^*$
- As a difference between the economical model, time proceeds in a fixed set of periods $t = 0, 1, 2$

The Audience size is fraction of the population that is participating in the social network and it could change dynamically over time:

- $t = 0$: initial audience size is z_0
- $t = 1$: people evaluate whether to participate based on the shared expectations that the audience size will be the same as what it was in the previous period, which means that $z_1 = g(z_0)$ since function g maps shared expectations to outcomes and so on and so forth

The population evaluate the benefit of participation as though the future will be the same as the present, which means that the number of people enrolled in this social will remain stable in time.

78.2 Analyzing the Dynamics



If we start from (z_0, z_0) that is the actual number of people that will join the site and hence we have (z_0, z_1) and this maps the enrollment to the site to another point (z_1, z_1) so there is an expectation that z_1 parts of the population will join the social. So on and so forth until when we reach (z_2, z_2) that is a **stable equilibrium point**. While (z_0, z_0) is an unstable equilibrium point, since it draws user away from the site if they get much farther from it.

This dynamics for updating the audience size illustrates how stable and unstable equilibria govern the outcomes:

- pushing population out of stable equilibria, **will bring it back**
- pushing population out of unstable equilibria, **will make it flow away**

79 INDUSTRIES WITH NETWORK GOODS

We want to understand how an industry that works with network effects might exploit the existence of stable and unstable points

79.1 Lowering The Production Cost

A new product is introduced with a high cost of production and the equilibrium is at audience size = 0, if the cost falls because we were able to lower the production cost then we might have equilibrium, but now we need to convince a group of customers to reach one of the equilibrium points and this would be much easier if the cost continues to decrease since the first critical point decreases, and sales grows rapidly to the stable point.

79.2 Marketing a Product with Network Effects

The marketing of a new product will not succeed unless you get past the tipping point (z'), since before this point everything will fail. The industry needs to convince a large initial group to adopt the product before others will be willing to buy it with some tactics:

- set an initial low, introductory price for the good, perhaps even for free; if the product catches on, then the firm can raise the price and perhaps make enough profit to overcome the initial losses
- alternatively, identify "influencers", and convince them to adopt the good and this involves network effects, but not at population level.

79.3 Social Optimality with Network Effects

We know that equilibrium in a market with no network effects is also social optimal, but with network effects the equilibria are typically not optimal since if we are at equilibrium with audience size z , we have that for these consumers, the addition of other purchases would make their benefit increase (increasing social welfare), but no one else is willing to purchase

A general principle is that markets with network effects typically provide less of the good than is socially optimal.

79.4 Competition

If different firms develop competing new products, each of which has its own network effects, it is likely that one product will dominate the market with network effects. The product that is the first to get over its own tipping point attracts many customers and this may make the competing product less attractive

80 MIXING INDIVIDUAL AND POPULATION LEVEL EFFECTS

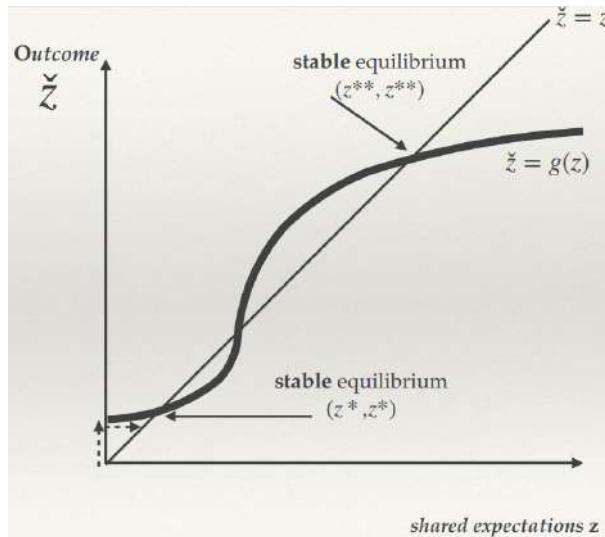
We have assumed that $f(0) = 0$ which means that if there is no one using the product, $f(z)$ is zero and hence it has no benefit. Now we want to analyze a case where a product has some value to a person even when they are the first purchaser, and this value increases as more people buy it.

We include both individual and population level effects in this model such as:

- $f(0) > 0$
- $f(z)$ increases with z

We will focus on the idea that small changes in the properties of the market can cause enormous changes in the size of the equilibrium audience that is reached, starting from zero.

80.1 Growing an Audience from Zero

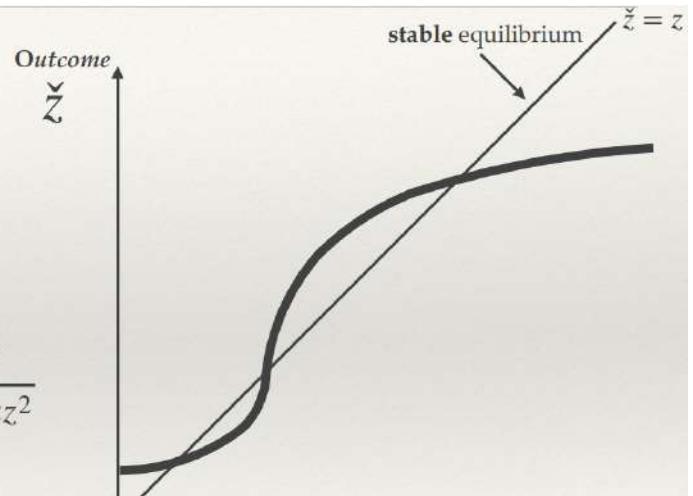


When $f(0) > 0$, an audience size of zero is no longer an equilibrium (when $p^* < 1$), and if the product starts at audience of size $z_0 = 0$ then the size increases up to the first point of the equilibrium.

z^* is a bottleneck since starting from an audience size of 0, we do not reach the more desirable equilibrium point z^{**} . The problem is that we are trapped in a stable equilibrium since here we don't have a tipping point.

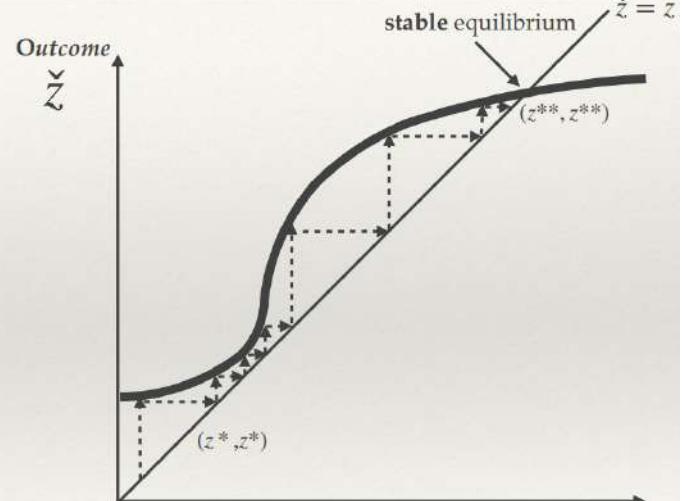
80.2 Bottlenecks and Large Changes

- ❖ The firm marketing the product aims to get z^{**}
- ❖ Suppose that the firm is able to lower the price p^* slightly, to some new value $q^* < p^*$
- ❖ We get a new function: $h(z) = 1 - \frac{q^*}{1 + az^2}$
- ❖ The curve moves upward until it no longer crosses the line $z = \check{z}$ at z^*



shared expectations z

- ❖ This is a totally different dynamic process
- ❖ As soon as the curve lifts enough that it no longer crosses $\check{z} = z$ near (z^*, z^*) , the audience sizes starting from zero increases continuously dramatically: from a value near z^* to a much higher value near z^{**}
- ❖ The bottleneck has opened into a narrow passageway



shared expectations z

81 CONSEQUENCES OF SMALL CHANGES

In models with network effects, small changes in market conditions can have strong and discontinuous effects on the outcome. When the price is p^* , we reach a small group of the most enthusiastic consumers, but it fails to get a much broader set of people, and when the price is slightly lowered to q^* , the product has the chance to become mainstream, pushing up the audience size. Hence we can go from a niche product to making it viral by just lowering the price slightly.

82 POWER LAWS AND RICH GET RICHER PHENOMENA

83 POPULARITY AS A NETWORK PHENOMENON

83.1 Popularity, Heterogeneity and Networks (Recap)

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

We know that real networks are actually **heterogeneous**, and that means that if we plot the degree distribution of the network and the probability of finding a node with that degree, we get in a log-log scale something as a lowering straight line. This is just a mathematical representation that is also linked with the presence of **hubs**, that are networks with high degree. The in-links between nodes can be used as a measure of popularity. The heterogeneity parameter κ is a measure of distribution's broadness and is the second moment of the distribution divided by the square of the first moment, which is the average. Larger the heterogeneous parameter the broader the distribution, when the parameter is close to one we can say that the distribution is quite heterogeneous

83.2 Case Study: The Web

The web is a good case study since the network that represents the web is a large scale network and we can retrieve samples of the web from many different sources, and hence the web is a good case study to show the **imbalances and inequalities of a network**. By analyzing the web we can observe that:

- almost everyone is popular for very few people, like for my family
- very few people achieve high popularity, like the actors or singers
- very very few people achieve global popularity, like maybe the president of the usa that is known world wide

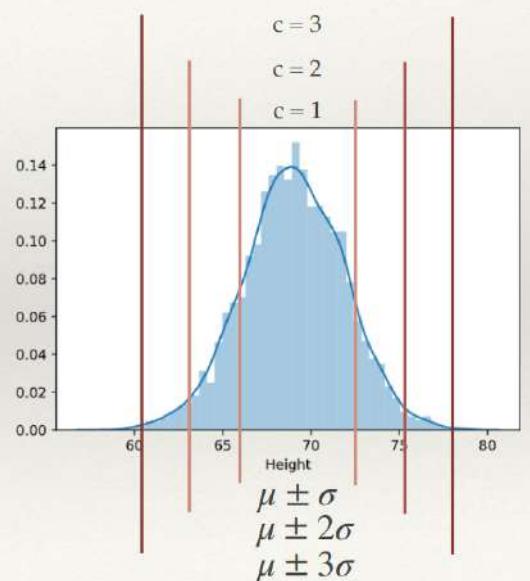
Hubs at a small scale of the network disappear when we look the whole network itself, and we can see that there are few hubs that have a global meaning and popularity. The property of finding hubs at every scale is **self similar because it is independent from the scale we use to analyze the network**, and we want to understand if this phenomenon is intrinsic and correlated to the idea of popularity itself, and hence we need to define a popularity scale.

83.3 Looking For a Popularity Scale

We need this scale to measure popularity and according this measure we can understand and define if a certain person is popular or not. We could make an example using the height of the human beings to determine if someone is tall, short or average height and to define these differences between the heights we need a scale defined on the heights like above 170 you are average, above 190 you are tall etc.. and we would need such a measure but in popularity.

In terms of networks we want to understand as a function of k , what fraction of web pages has k links to other pages. Of course the larger k is the larger the popularity is, and the fraction of web pages that have k links could be found out and understood using a probability distribution just as we could do for the heights of human people. The **mean** defines a scale of the population since and it is quite good for estimations and predictions.

- ❖ If we look at people's heights distributions
- ❖ mean: $\mu = 69.03$ (feet)
- ❖ std: $\sigma = 2.86$
- ❖ The prob. of observing a value that exceeds the mean by more than c times the standard deviation decreases exponentially in c
 - ❖ amazingly high persons are very unlikely



The standard deviation allows us to know that the probability of observing a value that exceeds the mean by more than c times the standard deviation decreases exponentially in c , so if we just consider $c = 1$ and we consider the interval of the mean + and - the standard deviation, we find out that a lot of the observed heights fall inside this interval. If we widen the interval that we consider we can get a larger fraction of the population, and going on with this process we can end up containing all the members of the study inside of the interval, and by doing so we can find out that the extremes of the graph we plot (extremely low persons and extremely high persons) are very unlikely to be found.

84 POWER LAWS

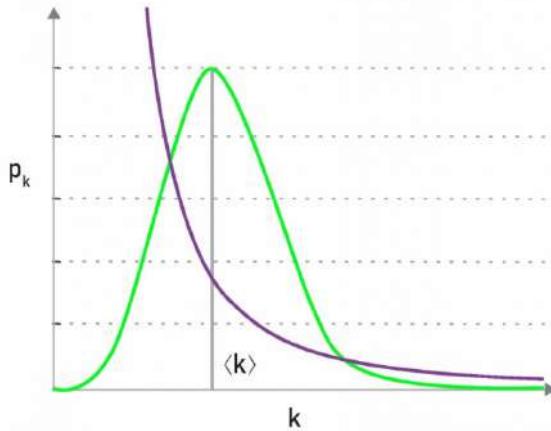
84.1 Empirical Findings for The Web – Power Law

- ◊ The fraction of web page that have k in-links is: $f(k) \approx \frac{1}{k^c} = k^{-c}$
 - ◊ $c = 2.1$
 - ◊ in other networks $2 < c < 3$ very often
- ◊ $f(k) = ak^{-c}$: **power law distribution**
- ◊ you can calculate that in the $2 < c < 3$ regime we have that when $N \rightarrow \infty$ then $\langle k^2 \rangle \rightarrow \infty$
 - ◊ **scale free networks**

For the web we get that the fraction of web page that have k in-links ($f(k)$) is the inverse of the degree up to c ($1/k^c$), that we can rewrite as k^{-c} , and it has been found that for the web c is 2.1, while in other networks it is between 2 and 3, and hence this finding was very surprising for the sciences. This function $f(x) = ak^{-c}$ is called a **Power Law Distribution**, and in the correct definition we apply a coefficient a before the function it self. If we compute the power law and we have that c lays between 2 and 3 and N tends to infinity, we find out that the second moment of the distribution $\langle k^2 \rangle$ will tend to infinity as well, this means that we have an infinite variance that makes, and a network that shows such a distribution is called a **Scale Free Network**, since they lack a scale.

When we deal with distributions that represent a bell shaped graph, we can always use the mean as a scale of the distribution, but with the power laws we achieve an infinite variance and this means that we cannot use the mean (or average) as a scale.

84.2 The Lack of a Scale



Power laws in the scale free regime is an unbounded distribution when the number of nodes N tends to infinity, so the population is infinite, but of course in reality we never deal with infinite populations so even if the variance is infinite when N tends to infinite, and hence the standard deviation is infinite, we can always calculate the mean and the variance. But in this case we have an explanation to why we use the heterogeneity parameter as the fraction of the variance divided by the power of the mean, since if the variance in our empirical data has been calculated, it would be so big compared to the mean that it will make the scale very lowly informative. The main consequence is that popular web pages are more common than we would expect with a normal distribution.

We can see in a graph that in our population nodes with a very small degree k can be over represented in respect to a normal distribution, and we can graphically see that the average of such a distribution is lowly informative, but if we tend to go to the right of the x axis we can easily understand that even pages or nodes in a network with high degree are much more common then what a normal distribution would lead us to think. Even if it is almost impossible to find an out of scale node following a certain distribution, the probability to find an out of scale node in a power node distribution is not neglectable.

84.3 Are Scale Free Networks Rare or Frequent?

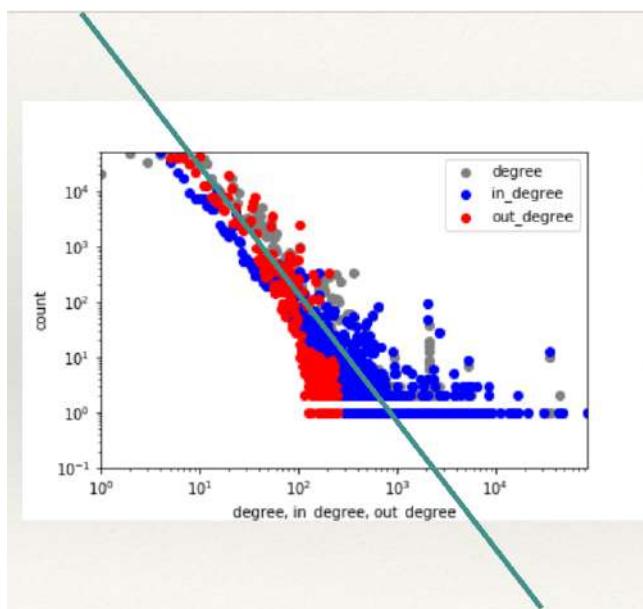
There is a long standing debate in the scientific communities if scale free networks are rare or frequent, and Barabasi wrote a book saying that scale free networks are very common while other scientists sustain that scale free networks are just theoretical frameworks and that they are not that common. The mutual agreement on this debate is that we can find scale free networks or power laws in empirical data according to the severity of the test, if we use a strong statistical test it is likely that we won't find a power law, but we actually don't really care, because for many practical reasons it is not important to know if a network is a scale free network since checking heterogeneity is often enough to distrust the average degree as a good estimator, and this brings all the other consequences like friendship paradox, robustness etc don't change if the network is a scale free network or if it can be fit with precision with a power law, but we want only to check the heterogeneity. But understanding some characteristics of power laws is important as much as looking for an underlying process that explains the emergence of hubs. Hence we don't care if these networks are realistic, but they could explain some aspects of reality in a simple manner.

84.4 Power Laws

◊ $f(k) \approx \frac{1}{k^c} = k^{-c}$ decreases much more slowly as k increases

Pages with very large k are much more common than expected with the normal distribution, and we need to abandon the idea of using a normal distribution because the emergence of hubs when we deal with power laws is quite likely. This can be observed empirically in many domains where we can find data that has a better explanation with power laws than it would with normal distributions, such as earthquakes with large magnitudes are very rare while earthquakes with low magnitude are very frequent every day.

84.5 Fitting with a Straight Line



- ◊ approximations of power laws are very common
 - ◊ $f(k) = ak^{-c}$ for some constants a and c
 - $\log f(k) = \log(ak^{-c})$
 - $\log f(k) = \log a - c \log k$
 - $y = \log a - cx$
- in a log-log plot:
 $\log a$ is the **intercept**
 $-c$ is the **slope**

We can approximate power laws with a straight line, if we apply the log to both sides of the equation a make some algebra we can draw a function in the graph that is the log of $a - cx$ and this creates a straight line in a graph where $\log a$ is the intercept, so the tilt degree of the line, while $-c$ is the slope that indicates where the line intersects the y axis.

84.6 Why Do Hubs Emerge?

If we accept that power laws represent many phenomena like Barabasi says, we want to understand how this happens. If we focus on the straight line that we create in the log log plot, it is agreeable to say that it is an order generated from chaos. Apparently the network is created randomly, but we observe at an higher scale is this straight clear line and hence there must be an underlying process that allows us to compute and plot such a straight line.

85 RICH GET RICHER MODELS

From cascade knowledge we can assume that people have the tendency to copy the decision of people who acted before them, and we assume that this behavior characterizes the creation of the network, in such a way:

- 1. Nodes are created in a sequence from 1 to N
- 2. For each node j that joins the network, we repeat:
 - (a) **with probability p** , page i is selected uniformly at random, and a link (j, i) is created
 - (b) **with probability $1-p$** , page i is selected uniformly at random, l is the page i is connected to, then a link (j, l) is created. Here we are copying the behavior of i and we are considering connecting to the pages that i is connected to, which is l and hence we create a link (l, j)

We imagine of having a certain threshold of the probability p (0.5) and we introduce a first node in the network. Then we introduce a second node that has a random number that we suppose is lower than the threshold, and we create a link from the second node to another node. If we introduce a third node we repeat this step and it has a lower random number compared to p and we connect it again to another node in the network. If we introduce a node that has a random number greater than the threshold, we pick up a random node and we copy its behavior, hence we connect the new random node to the nodes whose the node we are copying is connected.

The rich-get-richer model is a generalization of the preferential attachment model by Barabasi and Albert, since:

- **with probability p** the selection of the end-point is random because if p is very high we would generate a random graph
- **with probability $1-p$** we can prove that the second step of the rich get richer model (b) is equivalent to saying that the page j chooses a page with probability proportional to l 's current number of in-links, which is just the preferential attachment.

86 THE UNPREDICTABILITY OF THE RICH GET RICHER EFFECTS – IMPACTS OF POWER LAW

86.1 The Fragility of Popularity

Let's accept the idea that power laws are produced by feedback effects, we start copying the others behavior and doing this we will make the degree of other nodes higher, and this increase will increase as well the probability of selecting this node again, **the increasing popularity of a certain node will make it even more popular**. As an example we can say that a certain song is not maybe that good but if everyone listens to it then maybe we also would give it a try. The initial stages of the process that gives rise to the popularity of a node is a fragile thing and we want to focus on the notion of cultural market, hence we want to understand if we can predict the popularity of a certain thing like a song or a movie. We can expect initial fluctuations of certain movies, songs, or nodes in a network but this introduces quite a bit of **unpredictability** on a single node.

We can predict that a power law can emerge after a while, and so we can predict that we will have hubs. If we have a very large population we can predict that some one in this population will be hub and be very famous and successful, **but we want to understand which**

hub will be successful and this is quite more difficult to do since predicting the success of an individual item is not like predicting that some individual will have global success.

86.2 The Music Lab Experiment

MusicLab was a site where you could listen to songs, and download your favorites and they released a lot of songs from unknown artists and the songs had very different qualities among them, some very good songs and some very bad. The users of the sites were randomly assigned to different sessions and in some cases they didn't know how many people downloaded a certain song, and in some other cases they did know. The download count was used as a measure of popularity, since the higher the number of downloads, the higher is the popularity of that song, and from every category we can make a ranking of the songs.

At the end very good songs did not end up at the bottom and very bad songs did not end up at the top, while the other songs placed in mixed positions. In some sessions the order was established by means of popularity, which means that the participants of some sessions were able to find at the top rankings the most successful songs and hence **social influence was quite relevant at the end of the process** and this means that the initial fluctuations made an influence on the later participants since they found the songs at the top of the rankings and started listening and downloading them, and hence **initial fluctuations are quite unpredictable**.

This is very similar to information cascades, if we don't know what the earlier participants selected we cannot be influenced and we will make our decision on our own, while if we see some rankings we will influence our decision based on the rankings and the previous decisions.

87 THE LONG TAIL

The long tail is another impact of the power law. Popularity can be characterized by power laws which means that small sets of items can be quite popular but do we want to put our money on these niches used by a small amount of people or on hits used by every one?

Anderson proposed to do not focus on hits, but try to estimate the market sales of all the niches. We know that certain items, songs, etc.. are quite popular, and we also know that the majority of this things will be sold to fewer and fewer people. Almost intuitively we would need to focus on hits that would be sold to many people.

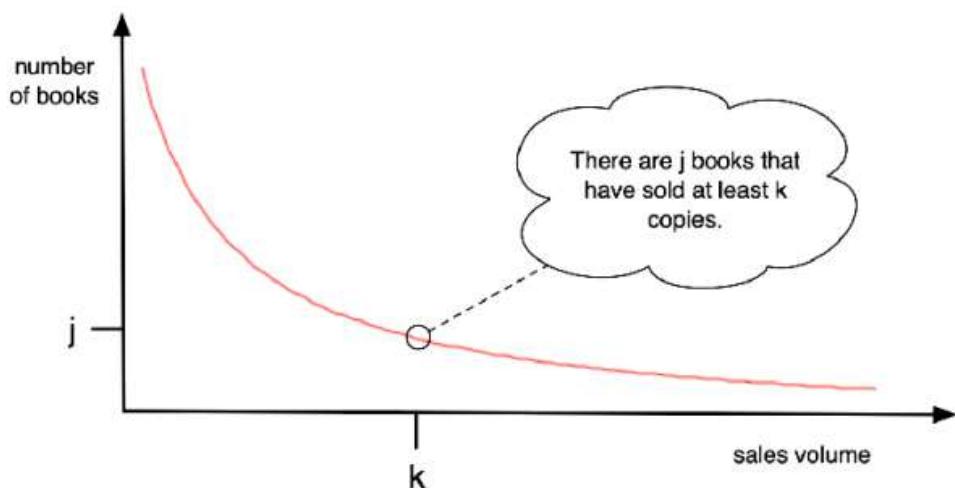
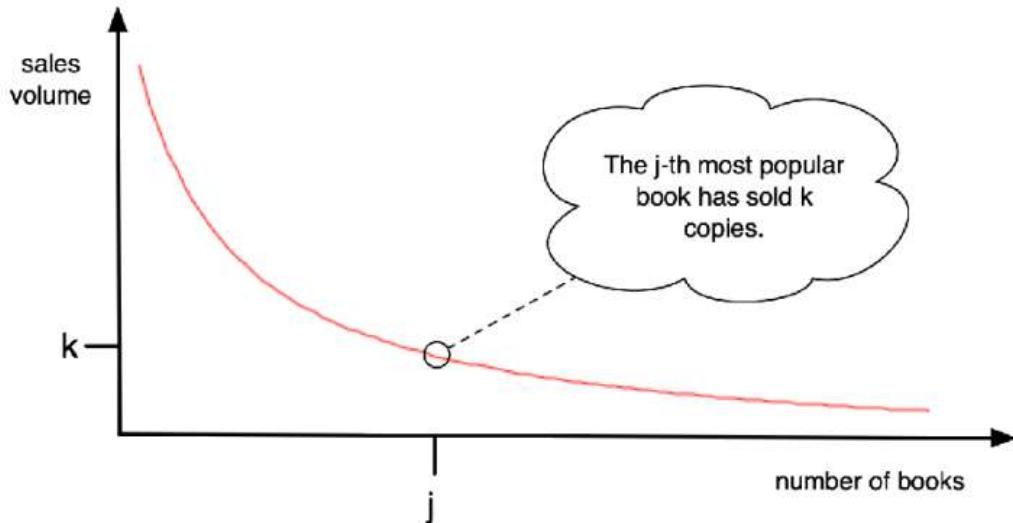


Figure 18.3: The distribution of popularity: how many items have sold at least k copies?

If we plot on a graph the numbers of items (j) on the y and the number of sales (k) on x we could have an heavy tail, and by doing this we can understand how many copies of that thing we have sold by finding the intersection of j and k in the graph. **So we want to understand as a function of k , what fraction of items have a popularity that is exactly k ,** and this is what the representation of a power law can help us to understand and represent. In the narrow part of the heavy tail on the right we know that very few items will be sold a lot of times, while on the other side we know that a lot of items will be sold very rarely, and if we switch the axes we can see this in another way.



By switching the axes we want to understand, as a function of j , what number of items have popularity at least k . In this plot we have a rank where we can see which item is the best seller on the left part of the graph. If we look at the shape of the graph we can compute the area underlying certain parts of the graph and if we understand that the area in the narrow side of the tail is smaller than the fatter side of the graph then we can consider the hits, while if it is the opposite it is better to have a catalog of niche items.

This type of plot is known as a Zipf's plot

87.1 Zipf's Law

Zipf's law refers to the size k of an occurrence of an event relative to its rank j , and it states that the size of the j -th largest occurrence of the event is inversely proportional to its rank. This means that k is approximately j^{-b} where b is close to 1.

87.2 Pareto's Law

Instead of asking which was the j -th largest income, he asked how many people have an income greater than j , and his law is a cumulative probability distribution: $P(K > k)$ is approximately k^{-y}

Zipf's, Pareto's and the Power Law are all connected to each other since we can prove that the exponent c of a power law is equal to $1 + y$ where y is the exponent of the Pareto's Law and is equal to $1 / b$, so if we plot a cumulative probability distribution from a Power law we would get a Pareto's law, and Pareto's exponent y is equal to the inverse of Zipf's exponent.

88 THE EFFECT OF SEARCH TOOLS AND RECOMMENDATION SYSTEMS

Search tools make the rich-get-richer dynamics more evident, if a lot of people link to a certain page, the page rank of that page increases or decreases and it will be quite likely to find this page in the top ranks of a google search for example. But if we find this page very easily we can say that it is an amplification of the preferential attachment since it is in the top rank of the result set and hence it will be even more likely to be connected to this page, and we can say that search engines boost the rich get richer effect.

Other aspects though can make this effect less extreme:

- different queries brings to different Search Engine results
- targeted and personalized search makes unpopular items ranked first but can be highly relevant for a particular individual
- recommendation system's **serendipity**, which is a measure that will help to make very suggestive suggestions exploiting the long tail argument

Search engines and tools need to take into consideration all these things to make an effective information retrieval algorithm or tool.

89 ANALYSIS OF THE RICH GET RICHER MODEL

The main problem here is that we have large scale networks (big data) with a lot of edges and we start to understand some structural characteristics by analyzing the structure. We start from understanding the average degree and the degree distribution. Degree distribution is a key factor in order to understand different characteristics of our network and the domain that the network is describing, **and one characteristic of real large scale networks is that they are heterogeneous and hence we have the emergence of hubs, which are nodes that are highly connected with respect to other nodes of the network**, like individuals that are much popular than others.

This heterogeneity follows some rules that we cannot fit in a normal distribution and this is a problem since normal distributions are widely used in statistical analysis. If we have to deal with heterogeneous networks we cannot use this distributions and the probability of finding some nodes whose degree (or in degree in directed networks) is much larger than the scale of the mean of the distribution and it is not neglectable. This isn't very common but it can still happen.

If we cannot use the normal distribution we need a mathematical rule that helps us to characterize this distributions and we want to use the **power law**.

Power law can be defined as a function of value x that is proportional to x up to a negative small exponent c , that usually is between 2 and 3. **We are lucky because when dealing with heterogeneous networks we cannot use gaussian distributions or bell like distributions but power laws are a good candidate to fit empirical distributions that characterize the degree of a large scale real network.**

The main impact of not using bell like distributions is that we cannot use the average as an estimator to predict the performance in terms of popularity of a given node. But we can use a general law that is characterized by the power law.

If we are the managers of an actor or a singer we would like to have a way to predict the success of a song or a movie and bet on it, and with the power law we can characterize the distribution of popularity of a node in a large network, and hence we have a key to kinda understand success but we do not have a way to understand what scale a random node of a population could have, this means that at the beginning of the dynamical process that makes all the nodes grow in popularity, we do not have a way to understand in these early stages which of the nodes (individuals) will be successful.

We know that social influence, hence the tendency to conform, may have a significant impact on the way people decide to buy a certain product. The idea is that if we have the tendency to copy the others behavior we can design a process which is the rich get richer process. Once we let a certain individual join a population, this individual connects with an edge to a pre existing node, this process is biased since we have a probability p to decide if we assign an endpoint to this new comer or otherwise we can decide to select as an endpoint a node based proportionally on the number of links that it already has. There will be an higher probability to an already popular node to attract other nodes, and this is the preferential attachment strategy.

89.1 Objectives

We need to formalize certain aspects:

$f(k)$ is the fraction of nodes with k degree, we count all the nodes with a given degree and then normalize by the total number of the nodes we have

The goal is that we want to prove that this fraction of nodes $f(k)$ is proportional to a power law such as k^{-c} and we will prove this starting from the rich get richer process, **since if we have this process the emergence of a power law must be expected**. If we get the power law starting from the rich get richer model we have a valid explanation of the formation of heterogeneous distributions in our network.

89.2 Recap of The Rich Get Richer Process

- 1. Nodes are created in a sequence from 1 to N
- 2. For each node j that joins the network, we repeat:
 - (a) **with probability p** , page i is selected uniformly at random, and a link (j, i) is created
 - (b) **with probability $1-p$** , page i is selected uniformly at random, i is the page i is connected to, then a link (j, i) is created. Here we are copying the behavior of i and we are considering connecting to the pages that i is connected to, which is i and hence we create a link (i, j)

- ❖ $X_j(t)$ random variable

= # of links to j at a time step t

- ❖ $X_j(j) = 0$

$$\diamond X_j(t+1) = X_j(t) + \frac{p}{t} + \frac{(1-p)X_j(t)}{t}$$

expected change in $X_j(t)$

With $X_j(t)$ we want to predict the numbers of links at a given time t connected to j, so its a way to estimate the in degree of j at time t. When we add a node j to the network we add it at time j since we are going to add the nodes one after the other and hence $X_j(j) = 0$ since when j is inserted it has no incoming links. We can try to estimate X_j at time $t + 1$ as the value of X_j at time t plus the the expected change with the respect of the previous timestamps. The expected change is made of p / t which means that there is a probability p / t of randomly selecting a node, and the other component is the complementary probability multiplied by the expected number of links of j at the previous timestamp t divided by the total number of nodes t inserted up to that point.

89.3 The Deterministic Argument

We could just use a probabilistic approach to prove this, but we instead use a deterministic argument, which is a method preferred by physicians and mathematicians.

- ❖ Let's suppose that:

- ❖ time runs continuously from 0 to N

- ❖ $X_j(t)$ is a continuous function

- ❖ It is like we are ignoring probabilities, and our idealized physical system just starts from a set of initial conditions

$$\diamond X_j(t+1) = X_j(t) + \frac{p}{t} + \frac{(1-p)X_j(t)}{t} \implies \frac{dx_j}{dt} = \frac{p}{t} + \frac{(1-p)x_j}{t}$$

$$\diamond \text{let's set } q = 1 - p \implies \frac{dx_j}{dt} = \frac{p + qx_j}{t}$$

The good thing about this approach is that if we consider the expected value of X_j at time $t + 1$ as we defined before, if we consider the continuous assumption we can define the expected change in terms of derivatives of differential equations, and the expected change is defined in a continuous point that is X_j but just in a differential equation type.

$$\frac{dx_j}{dt} = \frac{p + qx_j}{t}$$

$$\Rightarrow \frac{1}{p + qx_j} \cdot \frac{dx_j}{dt} dt = \frac{1}{t} dt$$

integrating both sides:

$$\int \frac{1}{p + qx_j} \cdot dx_j = \int \frac{1}{t} dt$$

$$\Rightarrow q \left(\frac{\ln(p + qx_j)}{q} + c_1 \right) = q(\ln t + c_2)$$

$$\ln(p + qx_j) = q \ln t + c$$

let's set $A = e^c$

we can exponentiate both sides:

$$p + qx_j = At^q$$

$$\Rightarrow x_j(t) = \frac{1}{q}(At^q - p)$$

Recall initial condition: $X_j(j) = 0$

$$0 = X_j(j) = \frac{1}{q}(Aj^q - p)$$

$$\Rightarrow Aj^q - p = 0 \Rightarrow A = \frac{p}{j^q}$$

We use the differential equations and use some mathematical passages to solve it.



We can substitute: $A = \frac{p}{j^q}$ into $x_j(t) = \frac{1}{q}(At^q - p)$

$$x_j(t) = \frac{1}{q} \left(\frac{p}{j^q} t^q - p \right) = \frac{p}{q} \left[\left(\frac{t}{j} \right)^q - 1 \right]$$

So we solved the deterministic approximation:

$$x_j(t) = \frac{p}{q} \left[\left(\frac{t}{j} \right)^q - 1 \right] \text{ is a closed form expression for how each } x_j \text{ grows over time}$$

89.4 Identifying a Power Law in the Deterministic Approximation

We now want to identify in the closed form expression we just got where the power law is.

For a given value of k , and a time t , what fraction of all functions x_j satisfies $x_j \geq k$?

$$x_j(t) = \frac{p}{q} \left[\left(\frac{t}{j} \right)^q - 1 \right] \geq k$$

Let's rewrite it in terms of j :

$$\Rightarrow j \leq t \left(\frac{q}{p}k + 1 \right)^{-\frac{1}{q}}$$

rewriting the inequality in terms of j :

$$\begin{aligned} \frac{p}{q} \left[\left(\frac{t}{j} \right)^q - 1 \right] \geq k &\Rightarrow \left[\left(\frac{t}{j} \right)^q - 1 \right] \geq k \frac{q}{p} \\ \Rightarrow \frac{t^q}{j^q} \geq k \frac{q}{p} + 1 &\Rightarrow t^q \geq j^q \cdot \left(k \frac{q}{p} + 1 \right) \\ \Rightarrow j^q \leq t^q \left(\frac{q}{p}k + 1 \right) &\Rightarrow j \leq t \left(\frac{q}{p}k + 1 \right)^{-\frac{1}{q}} \end{aligned}$$

QUEL SEGNO DEL CAZZO VUOL DIRE PROPORTIONAL

$$j \leq t \left(\frac{q}{p}k + 1 \right)^{-\frac{1}{q}}$$

Out of all the functions x_1, x_2, \dots, x_t at time t , the fraction of values j that satisfies the above inequality is:

$$F(k) = \frac{1}{t} \cdot t \left(\frac{q}{p}k + 1 \right)^{-\frac{1}{q}} = \left(\frac{q}{p}k + 1 \right)^{-\frac{1}{q}}$$

\Rightarrow we have the "shape" of a **power law**:

$\left(\frac{q}{p}k + 1 \right)$ is proportional to k

$$F(k) \propto k^{-\frac{1}{q}}$$

$-\frac{1}{q}$ is a negative exponent

$F(x)$: fraction of nodes with **at least** in-degree k

but we aim at finding an approximation for

$f(k)$: fraction of nodes with **exactly** in-degree k

that means we can approximate $f(k)$ taking the derivative:

$$\begin{aligned} -\frac{dF}{dk} &= -\frac{d\left(\frac{q}{p}k + 1\right)^{-\frac{1}{q}}}{dk} \\ &= \frac{1}{q} \cdot \frac{q}{p} \cdot \left(\frac{q}{p}k + 1\right)^{-1-\frac{1}{q}} = \frac{1}{p} \cdot \left(\frac{q}{p}k + 1\right)^{-1-\frac{1}{q}} \propto k^{-\left(1+\frac{1}{q}\right)} \end{aligned}$$

89.5 Final Step

The deterministic approximation of the model predicts that:

$$f(k) \propto k^{-\left(1+\frac{1}{q}\right)}$$

that is a **power law with exponent**:

$$1 + \frac{1}{q} = 1 + \frac{1}{1-p}$$



89.6 Meaning of the 'c' Exponent in function of p

Let's study the behavior of the exponent:

$$\lim_{p \rightarrow 1} \left(1 + \frac{1}{1-p} \right) = \infty$$

the exponent is infinity when link formation is mainly governed by uniform random choice ($p \rightarrow 1$): very large numbers of in-degree are extremely rare

$$\lim_{p \rightarrow 0} \left(1 + \frac{1}{1-p} \right) = 2$$

the growth is mainly governed by the preferential attachment process. The power law's exponent decreases toward 2, allowing for nodes with very large in-degree

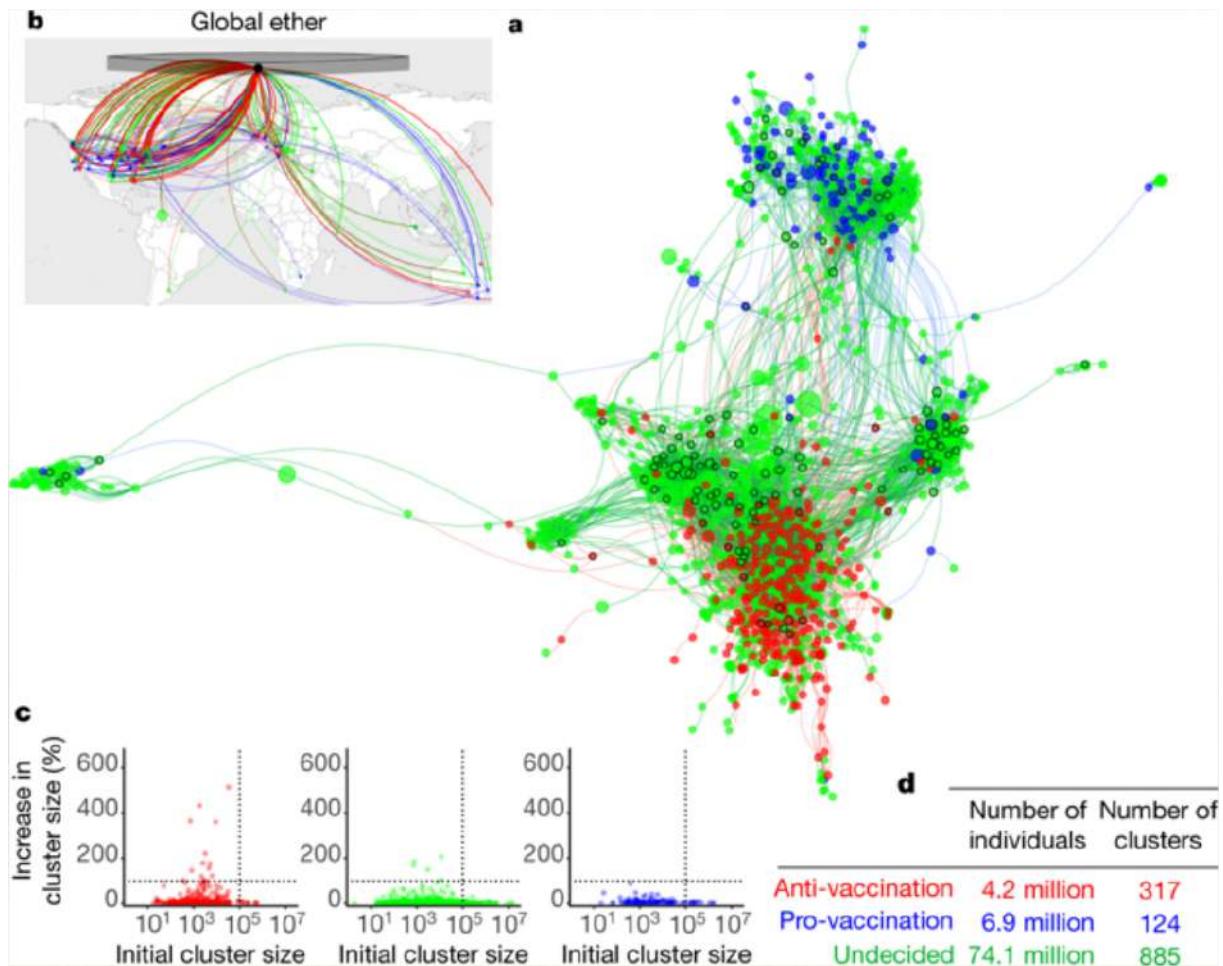
QUEL SEGNO DEL CAZZO VUOL DIRE PROPORTIONAL

89.7 Conclusion

Rich Get Richer processes explain the emergence of power laws and also exponents that in real scenarios are often slightly larger than 2 and this is another important signal to say that preferential attachments are an explanation of what happens in real networks.

90 CASCADE BEHAVIORS IN NETWORKS

There was an article between pro and anti vaccination views since it is a relevant topic nowadays due to covid. The authors of this paper analyzed a set of facebook pages and they found a polarization between the people that want to vaccinate (blu) and the people that don't want to vaccinate (red), while the majority of the people are undecided and leaning towards both ideas (green).



We can create clusters based on the ideas of the people regarding the vaccine. The argument of the authors was that the distribution of anti vaccination pages was more scattered in the facebook networks and if we run some models to understand the opinion shift towards anti or pro vaccination, the distribution of anti vaccination pages was more dangerous since it was more likely for an undecided user to be more influenced by an anti vaccination page.

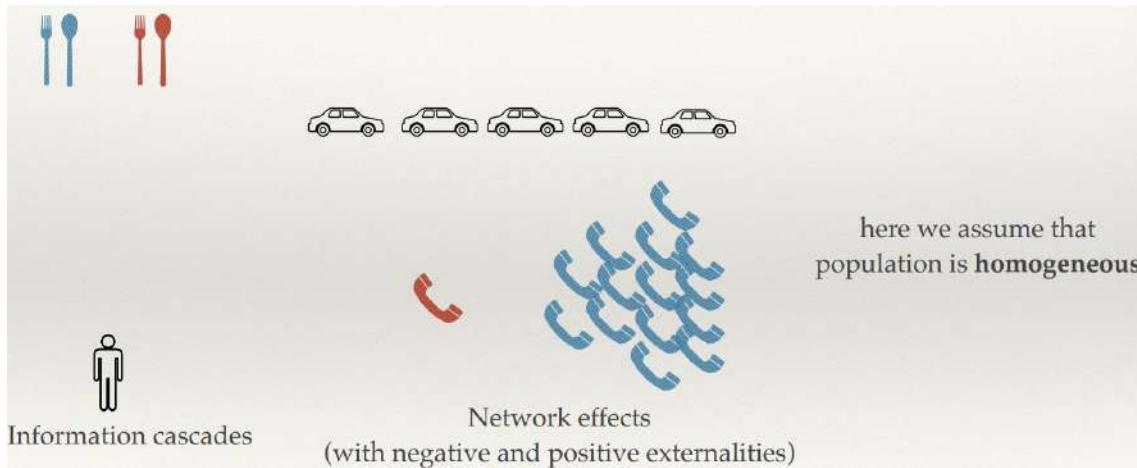
This is weird since the pro vaccination pages are much more then the anti vaccination pages, but they are concentrated in fewer clusters, while anti vax pages are scattered between much more clusters.

We want to understand the relationships between the cascades that we observe in such networks where we find different opinions and want to understand the formation and influence of this opinions, but also how opinions influence and modify the ideas and behavior of certain people.

We also want to understand how clusters in a network can create **barriers** for a given behavior or opinion allowing them not to spread, or in some cases they can act as **amplifiers** for ideas and behaviors and amplify the diffusion instead.

91 DIFFUSION IN NETWORKS

91.1 Individual's Choices Depending on Others

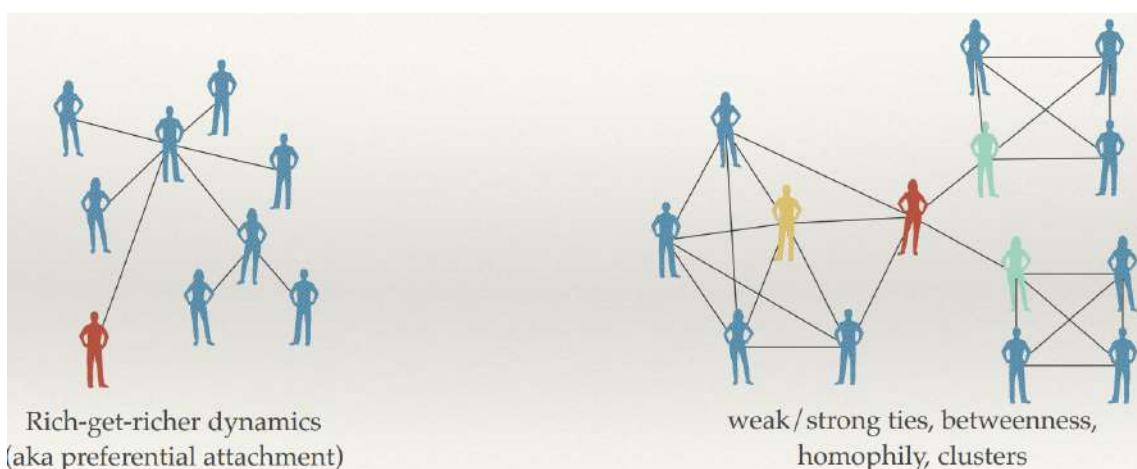


If we consider an example of where we want to go to eat dinner, we can observe the number of people in front of the restaurant and consider the herding effect, since after a queue is formed in front of a restaurant (even if its reviews are bad) we tend to decide and go to eat as well in that restaurant following the herd of people that wants to go to eat there. Same example could be made with the routes in a highway, if a lot of people drive on a certain route to avoid traffic jam, we tend to do the same. Same thing with a cellphone, if we are the only ones with a telephone we won't have many benefits, but if a big network of people joins the network we are much more likely to have benefits by using our own cellphone, and these are the network effects.

In these examples and situations we are assuming that the population is homogeneous, since we count the number of people that make a certain choice. We kinda estimate a number of people that are interested in eating at a certain place or taking a certain route, but there is no inner assumption about the structure of the network itself, but we can consider the structure of the network and hence consider heterogeneity as well.

91.2 What if we Consider Heterogeneity?

If we consider the network structure and hence heterogeneity, which means that some users in the network are more likely to be hyper connected with respect to the average from other people (as we know from power laws).



Rich get Richer dynamics (preferential attachment) are put into play, and there is a natural tendency to follow not only a crowd but to follow some nodes of the network more than certain others.

Structures of the network can be much more complicated since we can have different roles within like hubs central in a certain community or structural holes that connect two different communities that without them won't have a link, and this node can have an high betweenness and have a weak and strong tie link connection. Homophily plays a role as well into this, and we can say that the networks are not only created by rich get richer process, but they are shaped according an interest based on a selection or influence process between the nodes, and hence a certain network could have a lot of clusters.

91.3 The Diffusion of Innovations

We consider how new behaviors and opinions or even technologies spread from person to person through a social network, **as people influence their friends to adopt new ideas**, and we know that this problem of diffusion of innovation is studied by the informational cascades and networks effects. We must look and find **direct benefit effects** to understand how technologies spread, and what are the incentives to adopt a certain technology or behavior based on friends that already adopted those technologies or behaviors. To do this we need to take into consideration network effects on a local network level and we need a model to comprehend the spread of an innovation through a social network.

92 MODELING DIFFUSION THROUGH A NETWORK

92.1 Diffusion of a New Behavior

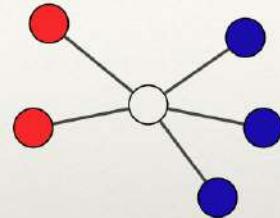
We are referring to the social contagion processes and hence **the assumption is that individuals make decisions based on the choices of their neighbors**, but the focus is not only the number of people that follow a certain group, but mainly on links and on local effects as of what is happening right now in our social circle.

92.2 A Network Coordination Game

We can try to model these effects as a coordination game, where two or more players will get an higher payoff if they coordinated. Each node has a choice, and players have an incentive to coordinate their behavior.

- ❖ It is natural to use a **coordination game**
 - ❖ each node has a choice between two possible behaviors, **A** and **B**
 - ❖ players have an incentive to adopt the same behavior

		w	
		A	B
v	A	a, a	0, 0
	B	0, 0	b, b



$$\begin{aligned}
 p & \text{ fraction of neighbors adopting A} \\
 1-p & \text{ fraction of neighbors adopting B} \\
 d & \text{ is the number of neighbors} \\
 \text{the node chooses A if } & pda \geq (1-p)db \\
 \Rightarrow p \geq & \frac{b}{a+b} = q
 \end{aligned}$$

If we are undecided on which behavior to adopt knowing that two friends adopt a certain behavior and three others another one, we can compute the preaction p of people that adopt a certain behavior A and $1 - p$ that is the other fraction that adopts behavior B. We find that d is the number of neighbors and hence we find that the undecided node chooses A if the $p*d*a$, where p is the number of people adopting A, d is the number of neighbors and a is the pay off, $\geq (1 - p)*d*b$, with this we find a **threshold** for p that is $p \geq b / a + b = q$. An undecided node will be activated if the fraction p of nodes that adopt A are greater or equal then the threshold that can be computed by the payoff matrix.

This threshold rule can help us to create a linear threshold model based on it. In order to adopt a certain behavior A or B we need to check if the fraction of activated nodes p is greater or equal then the certain threshold q .

92.3 Cascading Behaviors

In a certain network there could be two equilibria:

- one in which everyone adopts A
- another in which everyone adopts B

and we want to understand how easy it is to get one of these equilibria. Of course in both these equilibria not one node has an incentive to change behavior, unless he is obliged to. We also want to understand if there are other possible intermediate equilibria that are not this extreme, and how they could look.

We need to make certain assumptions:

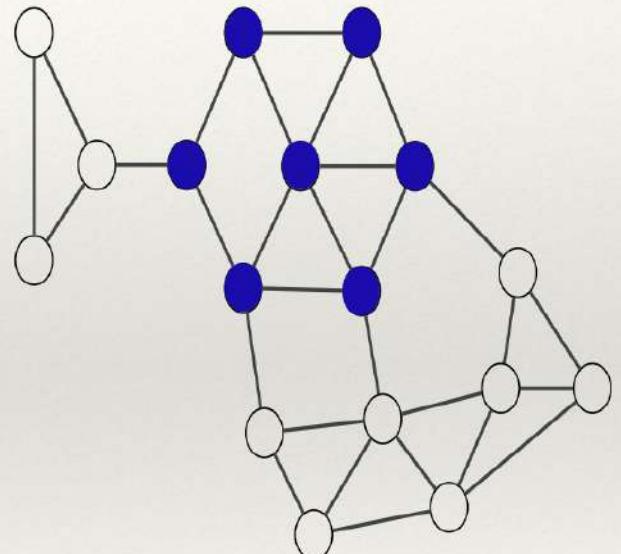
- everyone is using B at the beginning, or we could say that everyone is de-activated at the beginning
- S is small set of initial adopters of A, hence S is a set of activated nodes of initial adopters of A

We want to understand if the spread of A in the network will make everyone switch to the new behavior or technology or if the spread of the initial activated nodes will stop. The answer of course depends on the structure of the network, the size of S and the value of the threshold q .

We consider a set of initial adopters who start with a new behavior A, while every other node starts with behavior B. Nodes then repeatedly evaluate the decision to switch from B to A using a threshold of q . If the resulting cascade of adoptions of A eventually causes every node to switch from B to A, then we say that the set of initial adopters causes a complete cascade at threshold q .

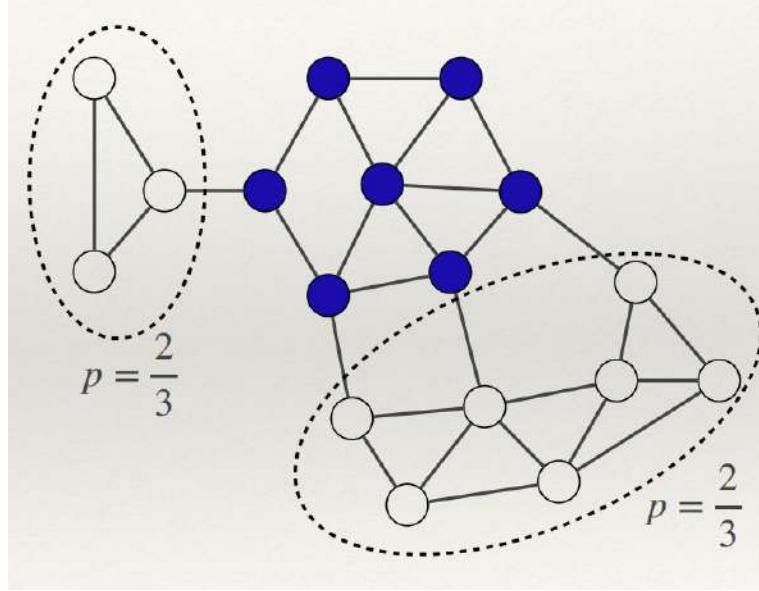
92.4 Monotonically Spreading

- ◊ claim: Adoption of A spreads monotonically
 - ◊ Nodes switch $B \rightarrow A$, but never back to B
- ◊ proof (by contradiction)
 - ◊ suppose some node switched back from $A \rightarrow B$, consider the first node (not in S) to do so (say at time t)
 - ◊ earlier at time t' ($t' < t$) the same node switched $B \rightarrow A$
 - ◊ So at time t' the node was above threshold q
 - ◊ But up to time t no node switched back to B , so the node could only have more neighbors who used A at time t compared to t'
 - ◊ there was no reason for that node to switch at the first place \Rightarrow contradiction!



93 CASCADES AND CLUSTERS

93.1 Stopping Cascades



Under this simple linear threshold model, cascades will spread in just a single direction and they can either get a full coverage of the network or stop.

We know that clusters have a major role in stopping cascades, but we want to know what else prevents cascades from spreading.

We know that homophily is a trigger to community formation and it can also serve as a barrier for diffusion, since it is hard for innovation to arrive from outside densely connected communities.

This intuition could be quantified by means of connections and links only in such a way: **cluster of density p is a set of nodes C where each node in the set has at least p fraction of edges in C .**

We have hence an **Internal Cohesion**, where each node in a cluster has a prescribed fraction of its friends residing in the cluster. Nodes in the same clusters do not necessarily have much in common since:

- Any network is a cluster of density $p = 1$
- The union of two clusters of density p is still a cluster of density p , even if the two sub clusters have no links

And in fact, clusters in networks can exist simultaneously at different scales.

93.2 Relationship Between Clusters and Cascades

We want to understand the relationship between clusters and cascades in terms of the threshold of our model and the density of the clusters.

The claim we want to prove is that **a cascade stops if and only if it runs into a dense cluster, and that would mean that clusters are natural obstacles to cascades**. In a more formalized re writing we say that S is the set of initial adopters of A , with a threshold of q , and divide the claim in two branches to prove:

- (i) If the remaining network contains a cluster of density $> (1 - q)$, then S cannot cause a complete cascade.
- (ii) if S fails to cause a cascade, then there is a cluster of density $> (1 - q)$ in the remaining network

If we prove this claim we find a strong relationship between clusters and cascades that can be analyzed and understood in terms of p and q , hence in terms of density and threshold.

93.3 Clusters Imply No Cascade - No Cascade Imply Clusters

Proof by contradiction:

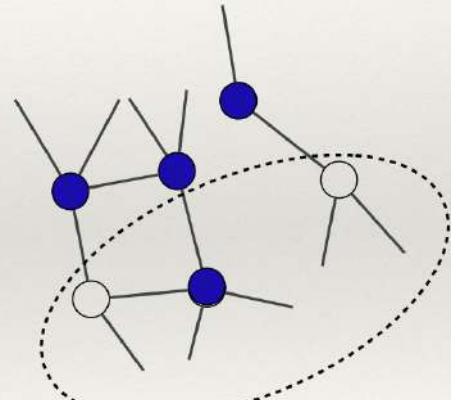
some node inside the cluster of density $> (1 - q)$ will adopt A at time t (the first to adopt in its cluster);

this means that the decision was made at time $(t - 1)$, when no other node in the cluster adopted A ;

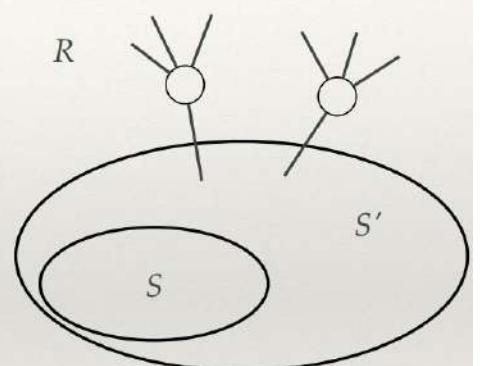
since the cluster has density $> (1 - q)$, less than a q fraction of nodes are outside the cluster;

nodes that could have adopted A were outside the cluster at time $(t - 1)$

it is impossible that at least q neighbors of the node adopted A :
contradiction



- ◊ Let S be the set of initial adopters of A
- ◊ The spreading process stops: let S' be the maximum set of nodes that switched to A
- ◊ Let R be the set of nodes still using B at the end of the process
- ◊ Pick one node in R : it does not want to switch to A
 - ◊ \Rightarrow the fraction of its friends using A is $< q$
 - ◊ \Rightarrow the fraction of its friends belonging to R is $> (1 - q)$
 - ◊ \Rightarrow this holds for every node in R
 - ◊ $\Rightarrow R$ is a cluster of density $> (1 - q)$



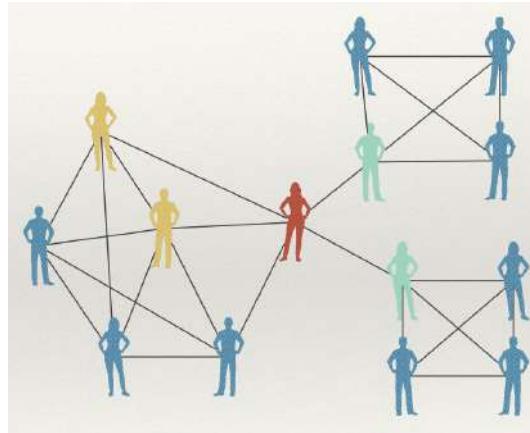
94 DIFFUSION, THRESHOLD AND THE ROLE OF WEAK TIES

94.1 Viral Marketing

We start from viral marketing and we say that tightly-knit communities in the network can work to stop the spread of an innovation and hence we have coexistence of both behaviors or technologies (apple vs android). In case of coexistence, if a firm can rise the quality of product A, than the spread of the new technology can produce a complete cascade. This means that the payoff of adopting A is greater and hence even a small fraction of people that have product A could incentivize me to buy A or adopt behavior A. **An alternative strategy** (when the firm cannot rise the quality of the product) is **trying to convince a small number of key people in the part of the network to still use B, but it is difficult to choose key people to trigger a viral marketing strategy**. There could be certain nodes in a social network that could be much more important and relevant to the spread of a marketing strategy compared to some other more anonymous nodes.

In a study they understood that there was a massive difference between learning about a new idea and actually deciding to adopt it, since the hybrid corn was adopted by farmers almost 10 years after they first were introduced to it. In many different situations people don't adopt a certain technology instantly, but they wait for some phenomenon that would trigger the usage of this technology, **it is like nodes at the border in a cluster, since they have been exposed to the idea, but still they decide not to adopt it**.

94.2 The Role of Weak Ties



Weak Ties connect nodes that are at the border of certain clusters. Threshold models, even if extremely simple, are good at highlighting some important implications of the strength of weak ties theory. Two certain weak ties, that are the border of two different clusters (verde chiaro), are exposed to some new cutting edge ideas, and their exposure is enough to make them aware of this idea, but not enough to make them adopt it and start spreading it.

Bridges and weak ties are great for spreading rumors or jokes across the network and the existence of a new idea, but not for diffusion of innovation or social mobilization, since talking about a certain product is almost costless, while buying it and spreading it is quite more expensive, if we don't have a payoff of course.

Strong Ties and people that are very central in their cluster/community (giallo) can have more significant role for others in the community to take actions and adopt a certain behavior or buy a certain product.

94.3 Simple vs Complex Contagion

The **Simple Contagion Process** says that the activation of a node depends on the threshold condition and if the threshold condition is satisfied or not. We basically use a linear threshold model based on the concept of peer pressure, on how many neighbors we need to influence other nodes.

In **The Complex Contagion Process** each new node that tries to activate a neighbor has greater influence than previous ones. If we receive the information that a new cutting edge product is released but we don't want to change our current device. If a certain friend tells me again how this new product is so much better than the previous, after a second time of hearing this information about this new cutting edge product, the probability of me buying this product is higher after hearing about it a second time. **Re-iteration of information about a certain product will make the probability of activation of a node, which means of me buying a certain product, higher.**

94.4 Centola's Experiment

Centola studied the spread of a health behavior through a network-embedded population by creating an Internet-based health community, comprising 1,528 participants recruited from health-interest World Wide Web sites. Each participant created an on-line profile, including an avatar, a username, and a set of health interests. They were then matched with other participants in the study, referred to as "health buddies", as members of an on-line health community. Participants made decisions about whether or not to adopt a health behavior based on the adoption patterns of their health buddies. Arriving participants were randomly assigned to one of two experimental conditions, which were a clustered lattice network and a random network, which were distinguished only by the topological structure of the social networks.

The results of the experiment brought an understanding between simple and complex contagion processes.

In the case of Simple contagion, a single contact with an "infected" individual can be sufficient to transmit the behavior. Under "the strength of weak ties" hypothesis, random networks with small world topology will spread a social behavior farther and more quickly than a network in which ties are highly clustered

While in the case of Complex contagion, the competing hypothesis argues that when behaviors require social reinforcement, a network with more clustering may be more advantageous, even if the network as a whole has a larger diameter.

We have a great result since we understand that different contagion processes can be stopped or amplified according to the existence and presence of clusters.

Centola investigated the effects of network structure on diffusion by studying the spread of health behavior through artificially structured online communities, and found out that the fraction of adoptions of new health behaviors was much higher in the clustered network compared to the random network, and hence he understood that clusters are a device that boost the spread and make it much higher. The reason for this is not that cluster stop the cascade from spreading, but it is that the perception of the risk from the point of view of an individual, decreases when they receive reinforcing signal, hence the second time an individual is exposed to the risk that adopting a certain behavior could reduce, the probability of adoption of a certain habit increases according to the **hazard ratio g**, that indicates that the probability of adoption increases by a factor of g for additional signals compared to the baseline hazard of

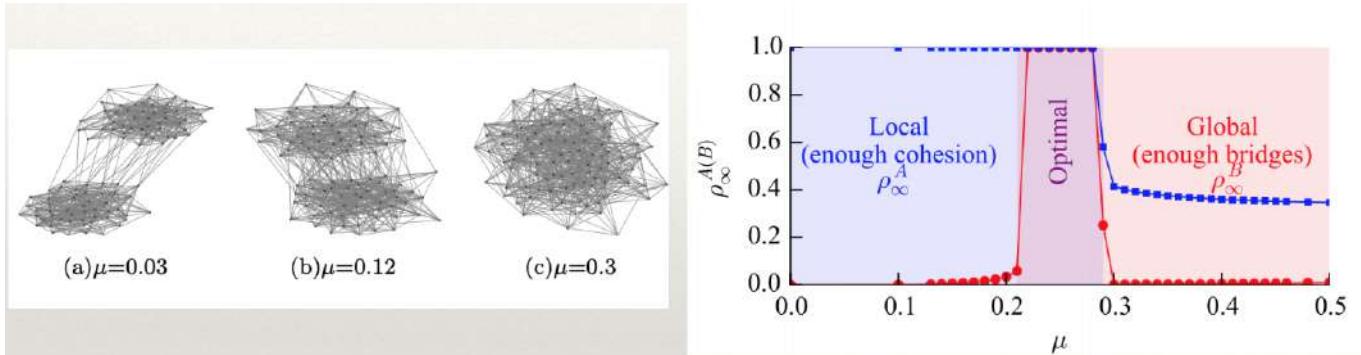
adoption from receiving a single signal.

Centola's Findings and Conclusions were:

- Individual adoption improved by reinforcing signals that come from clustered social ties
- This individual-level effect translates into a system-level phenomenon whereby large scale diffusion can reach more people, and spread more quickly, in clustered networks than in random networks.
- While locally clustered ties may be redundant for simple contagions, like information or disease, they can be highly efficient for promoting behavioral diffusion

Centola conjectures that public health interventions aimed at the spread of new healthy behaviors may do better to target clustered residential networks than the casual contact networks across which disease may spread very quickly, particularly if the behaviors to be diffused are highly complex.

94.5 Finding The Optimal Clustering For Spreading

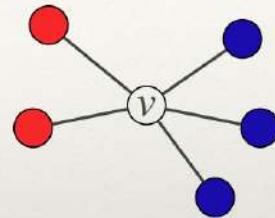


There is a modularity value (μ) where we can get a great speed up of diffusion inside a local community, while also having a global spread by growing bridges between communities. With modularity in this perfect range we can have spread in local and global community and hence have a great success.

95 EXTENSION OF THE BASIC CASCADE MODEL

- Let's suppose each person gives values to A and B subjectively

		w	
		A	B
v		a_v, a_w	$0, 0$
A	B	$0, 0$	b_v, b_w



p fraction of neighbors adopting A
 $1-p$ fraction of neighbors adopting B
 d is the number of neighbors
 the node chooses A if $pda_v \geq (1 - p)db_v$

$$\Rightarrow p \geq \frac{b_v}{a_v + b_v} = q_v$$

With these formulation we deal with networks that have different thresholds and not only one threshold.

Watts and Dodds found out that we need to take into account not just the power of influential nodes, but also the extent to which these influential nodes have access to easily influenceable people, which are people that need a low percentage of nodes adopting a certain behavior to adopt it themselves.

The notion of **blocking clusters** is now the set of nodes for which each node v has a fraction $> (1 - q_v)$ of its friends inside the set.

The notion of density becomes **heterogeneous** as well, since each node has a different requirement for the fraction of friends it needs to have in the cluster. Computation of these values will be harder of course.

96 KNOWLEDGE, THRESHOLDS AND COLLECTIVE ACTION

In many domains you need to integrate network effects at **both the population level and the local network level**, and we could have:

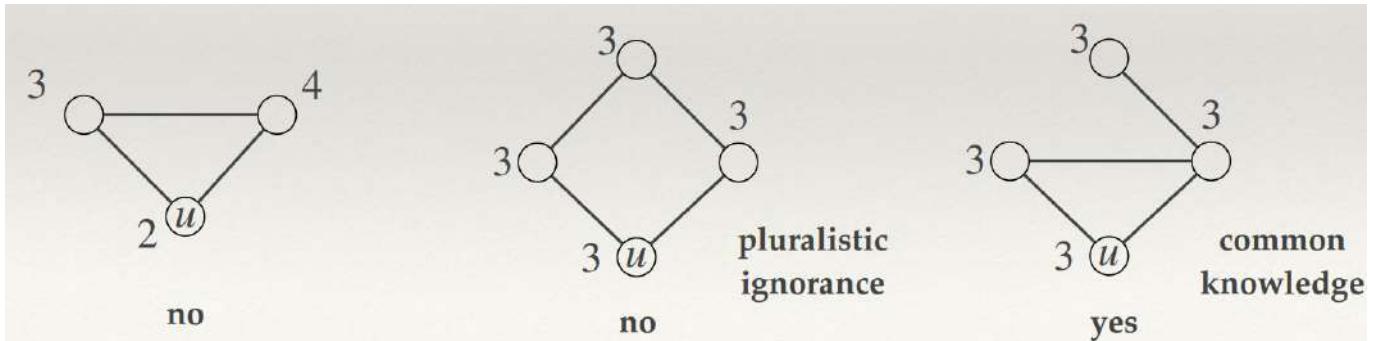
- collective actions and pluralistic ignorance
- a model for the effect of knowledge on collective actions
- common knowledge and social institutions

96.1 Collective Actions and Pluralistic Ignorance

The objective, ideally, is to organize a revolt under a repressive regime without getting caught and dying. If we apply **Collective Actions** we get benefits only if enough people participate, but you can talk to a very limited number of persons you trust. A problem is that we have no idea if this revolt is spreading enough and convincing enough people to take place into it, and this is **Pluralistic Ignorance**, when people have wildly erroneous estimates about the prevalence of certain opinions in the population at large.

96.2 A Model for The Effect of Knowledge on Collective Actions

Each individual as a personal **threshold of k** that says "I will show up for the protest if I am sure that at least k people in total will show up", and as an assumption we say that every node **knows the thresholds of all its neighbors in the network**.



96.3 Common Knowledge and Social Institutions

Some social institutions have the purpose of helping people to achieve common knowledge. Widely-publicized speech, article in a high circulation newspaper, have people receive the message and the information that a lot of other people are receiving the message. This was and is done widely in marketing by running commercials during a very popular event like the Super Bowl.

97 THE CASCADE CAPACITY

97.1 Infinite Networks

This is the theoretical framework that we use to explain how cascade could form and be fall and we can use these networks to understand how different network structures are more or less prone to cascades. **Cascade capacity** is the maximum threshold at which a small set of initial adopters can cause a complete cascade, and our objective is to find this cascade capacity. We consider infinite graphs with nodes that have finite numbers of neighbors.

Let's say that a finite set S of initial adopters of A causes a cascade in G with threshold q if eventually every node in G adopts A.

If we start with an **infinite path** and with one node in the set, his neighbors on the left and on the right will be convinced and the process repeats for all the new nodes in A.

If we start from an infinite grid the process is kinda the same, if we start from the middle of the grid all the nodes surrounding the certain node will fall into A and so on and so forth.

Even if we start with an infinite tree with the seed of A as the root of the tree, we can convince the neighbors, and the neighbors of the neighbors and etc..

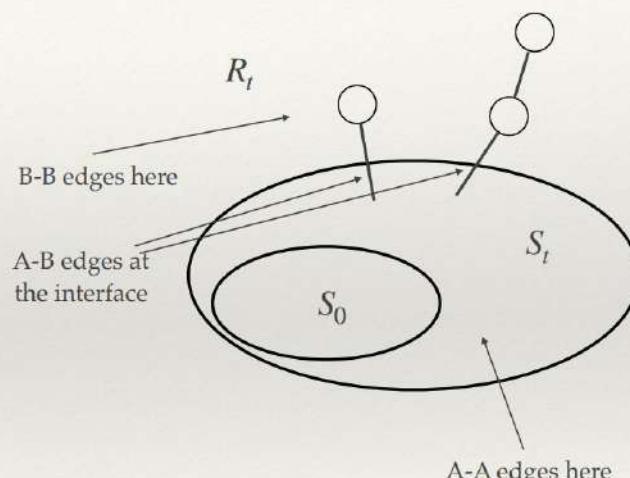
97.2 How Large Can the Cascade Capacity Be?

The claim is that there is no network of which the cascade capacity exceeds $1/2$.
We prove this by contradiction.

We suppose that there exists a network G with $q > 1/2$ and we want to find a contradiction by saying that nodes stop switching from B to A after a finite number of steps.

So we start by focusing on the interface

- ❖ S_0 : finite set of initial adopters of A
- ❖ S_t : set of adopters of A, potentially larger than S_{t-1}
 - ❖ size of the **interface** at time t : I_t
- ❖ We show that $I_t < I_{t+1}$
- ❖ Hence, if the size of the interface strictly decreases, the diffusion process will **terminate** after I_0 steps



97.3 The Size of the Interface

When $q > 1/2$ no finite set of nodes can cause a complete cascade in any network, and this threshold corresponds to a lower quality product or behavior and a person will switch to this device or adopt this behavior only if more than 50 percent of its friends have adopted that certain device or neighbor.

97.4 Another Extension of the Model

If we allow people to adopt both A and B we have a compatibility that is referred to as the bilingual option. But even if someone is allowed or can use both A and B it is very likely that it will evolve into using only one of its two options. This compatibility can be a strategy of a great firm to enter in a market and progressively cut out the competitor's product.

98 SMALL WORLD AND DECENTRALIZED SEARCH

98.1 Recap of Milgram's Experiment and Six Degree of Separation

Milgram selected in a random way some persons of the midwest and requested that they send a letter to someone that they don't know that lives in Massachusetts. All of these people know the name of the recipient, his profession and where he lives (but not the precise address). Each of the participants was requested to send their letter to a person that they knew that, according to their knowledge, had a higher chance to know the final recipient. The other person would then do the same until the letter reaches the final recipient. At the end of this experiment he observed that, in average, 6.5 steps were required to reach the final destination, hence we have the **Six Degrees of Separation**, that indicates that there are 6 steps between people that don't know each other. This was then called the **small world phenomenon** and indicates that the world is actually smaller than we may think.

This experiment showed amazing results since it showed that:

- We are connected, since even if we pick up two random persons in the world, they are connected by some path. This is surprising since there are a lot of persons in the world!
- We are close to each other, so not only we are connected but even close since we have between each all of 7 billions of us, there are 6 degrees of separation
- We are also collectively able to find short paths between all of us, so we have a way of not only finding a path, but even a short path in a decentralized way

98.2 Structure And Randomness

Connectivity and Small Distances can be explained with randomness and that links can be created randomly, and we can explain this with the Erdos Renyi Model:

- 1. Start with N nodes and zero links
- 2. Go over all pairs of nodes, for each pair of nodes i and j , generate a random number r between 0 and
 - If $r < p$, (where p is a selected threshold that is the probability) i and j get connected
 - If $r > p$, i and j remain disconnected

The idea behind this model is just to take a bunch of nodes, and for each pair of nodes we toss a biased coin and if we get head we establish a link, if it says tail we don't establish a link. The idea is that in this model edges are created randomly.

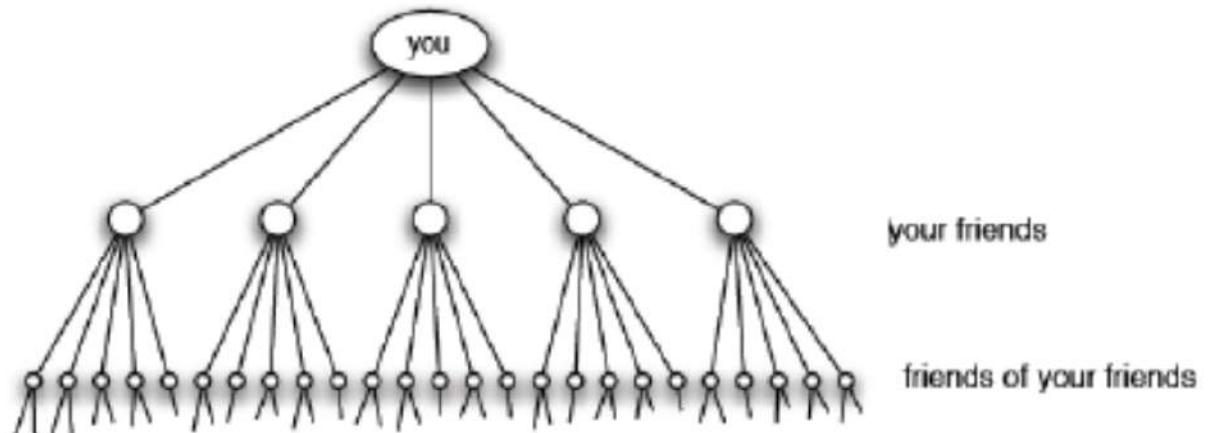
98.3 Evolution of Random Networks

- **If the probability p is 0**, that means that for all the pairs of nodes in the network there never is the chance to create a link, so at the end we have N components with one node each, so N singletons.
- **If the probability p is 1**, that means that every time the coin is tossed we form an edge, and hence all the links between the pairs of nodes will be created. If $p = 1$ we will have a complete network with N nodes.

As we add links to the network we can have two expectations:

- **Naive expectation:** the size of the largest component grows smoothly with the number of links
- **Wrong Expectation:** there is an abrupt increase for a given value of the link probability p

If we think about the naive expectation we expect a linear growth in the network and we see the emergence of a largest connected component, but to see that this emergence happens after the average degree surpasses 1 is counter intuitive because it is not dependent of the total number of nodes, but it depends only on the average degree. This can be generalized in networks of all sizes and orders.



If we build a network randomly we can think of our graph as a tree, since we are connected directly to our friends at distance 1, and the friends of our friends are at distance 2 etc..

Random Networks are good at predicting small world properties. We want to understand how many nodes are, on average, d steps away from any node in the network.

We can start from a node with a degree k , and we know that in a random network we can say that this degree k can be the average degree of the network. This means that:

- At distance $d = 1$ there are k nodes
- At distance $d = 2$ there are $k(k - 1)$ nodes
- ...
- At distance d there are $k(k - 1)^{dminus1}$ nodes

If k is not too small, the **total number of nodes within a distance d from a given node** is approximately:

$$N_d \sim k(k - 1)^{d-1} \sim k^d$$

Where there is an upper bound that is $k(k - 1)^{dminus1}$. At a given distance d we can find a total number of nodes that is approximately k^d .

98.4 How many steps does it take to cover the whole network?

$$\begin{aligned} N &\sim k^{d_{max}} \\ \log N &\sim d_{max} \log k \\ d_{max} &\sim \frac{\log N}{\log k} \end{aligned}$$

The total number of nodes N is approximately $k^{d_{max}}$, and if we apply the logarithmic operation we have that $\log N$ is approximately equal to d_{max} of $\log k$, which means that d_{max} is approximately equal to $\log N / \log k$ and this means that the diameter of the network grows like the logarithm of the network size.

We know that social networks have both short paths and high clustering, and we know that we can't predict high clustering in a network. We have **homophily and weak ties** since we have the emergence of clusters and shortcuts, and they provide these shortcuts between different social circles. But we know that random networks have small paths, but very low clustering (almost no triadic closure).

98.5 Watts-Strogatz Model Recap and Alternative Formulation

In order to understand the co-existence between the shortest path and high clustering coefficient we adopted the Watts Strogatz Model:

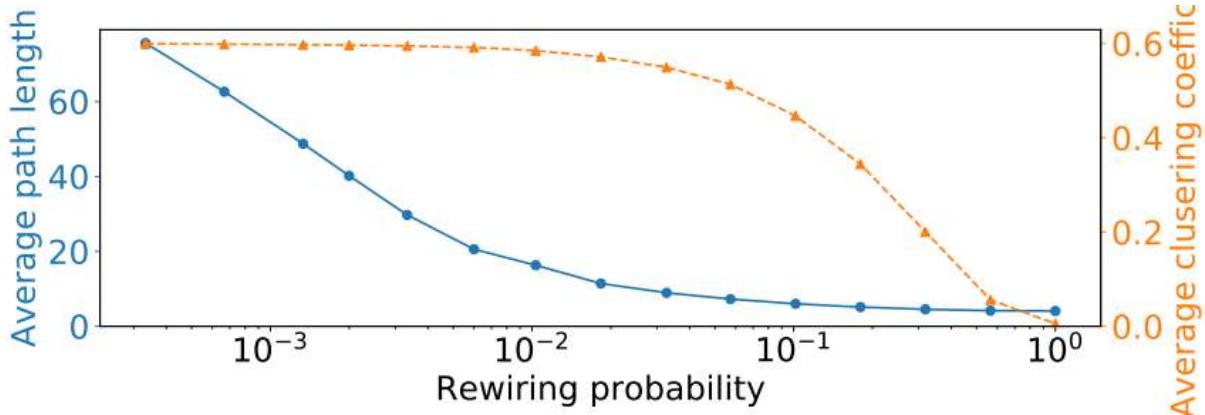
They started from a ring lattice, with N nodes that form it, with a degree k . With a probability p , the links are rewired randomly. The higher is p , the more the links are rewired, with $p = 1$ the whole edges are rewired.

The expected number of rewired links is $pL = pNk/2$, and we have that:

- If $p = 0$, no links are rewired and hence we have no change
- If p is small, few links are rewired and hence the average clustering coefficient stays approximately the same because very few triangles are destroyed, but distances shrink considerably
- If $p = 1$, all links are rewired and hence we have the opposite situation and the network becomes a random network, where we have short paths and short distances.

We need to find a probability p that allows us to keep an high clustering and a low average shortest path. Distances become short for low values of p , and the average clustering coefficient stays high up to large values of p .

There is a range of values of p where the average path length is short and the clustering coefficient is high.

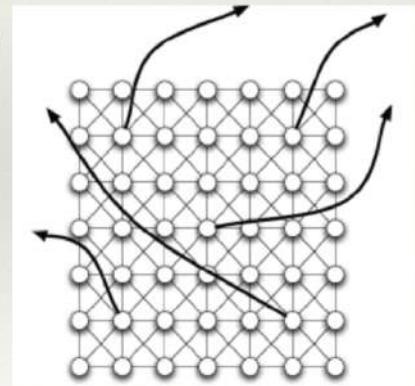


We can hence prove that high clustering coefficient can exist in random networks where the average path length is short, because of the presence of **shortcuts**. Every node has many connections with its inner social circle, but a node can have a weak tie with a node that is further away.

We can make an alternative formulation of the Watts Strogatz Model **in which we start with a square grid where every node is connected to his neighbors and the neighbors are connected with each other, and hence we know that in a grid the clustering coefficient is quite high**. Shortest paths aren't really that short since if we get the nodes at the opposite angles of the grid they are quite distant from each other.

The idea, even in this alternative formulation, is to rewire some of the links from a certain node to another in the grid by picking up the second endpoint of the new link randomly. In this alternative formulation we have an assumption that **each node has one random long range edge in the grid, which is a weak tie, but the other neighbors are all connected to each other and hence homophily is quite high**. We can also calculate the average clustering coefficient (C_i) that is close to 0.3 approximately.

- ❖ Start with a square grid
- ❖ Each node has 1 random long-range edge (a *weak tie*)
 - ❖ Other neighbors are connected each other (*homophily*)
- ❖ The diameter is $O(\log(N))$
 - ❖ (proof: by exercise)



If we asset the idea that each node has one random long range edge in the grid we can recalculate the diameter that is $O(\log(N))$.

We can finally confirm that high clustering coefficient and small distances coexist. In order to have small distances in a grid we just need to have a few random links which act as shortcuts to long distance nodes.

The Watts-Strogatz model is quite a powerful model because:

- provides insight on the interplay between clustering and the small-world property
- captures the structure of many realistic networks
- account for the high clustering in many networks
- does not predict the correct degree distribution, and hence we have no heterogeneity in this model
- does not enable navigation, hence we need other models to explain decentralized search.

99 DECENTRALIZED SEARCH

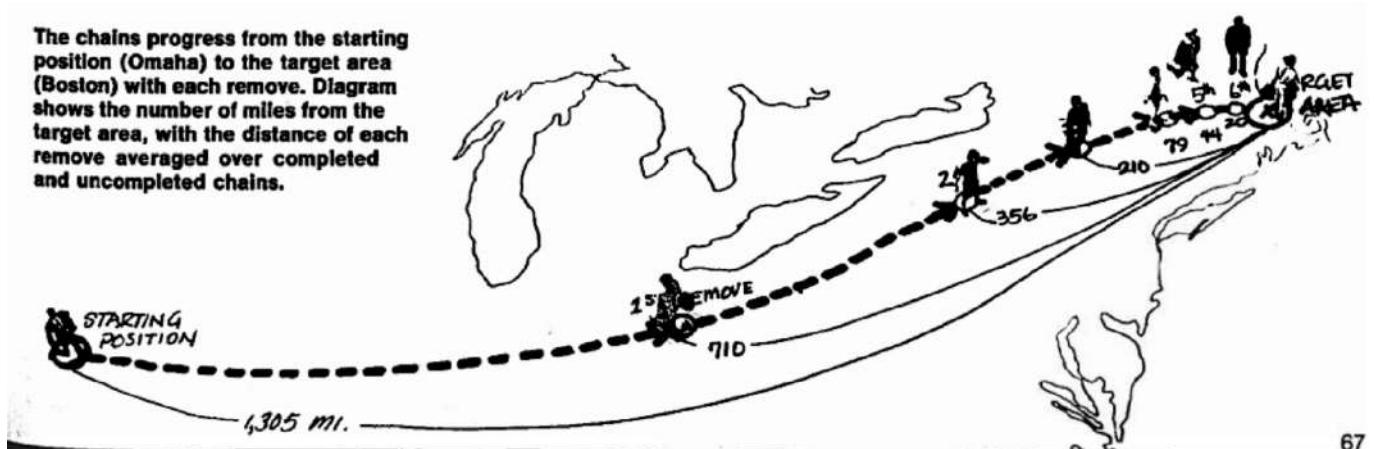
99.1 Searching for Short Paths in a Decentralize Way

Our claim is that we can find short paths to a given target with only local information, but we cannot use BFS since our favourite algorithms to find a shortest path work by assuming that we have a global view of a network. If we are individuals lost in a very large network and we don't know the knowledge on which are the links between the links that we don't know, and hence we can see only at a short distance close to our local view, and anything out of this range is unknown to us.

This is probably the most intriguing outcome of the Milgram's experiment because it is like the network contained some kind of 'gradient' that helped people to guide messages toward the target, so it is not surprising that the letters found the target and it is not surprising that these letters showed us that the distance between people is short.

We want to understand if we can we build a random network with some algorithm in which decentralized routing succeeds and what are the qualitative properties that are crucial for this success.

99.2 Distances From The Target are Constantly Reduced



67

Milgram proved that he was aware that the decentralized strategy adopted by some individuals that found the target in few steps was extraordinary since he found out that at the very beginning of the process individuals used to send the letter as far as possible from them, and the distance after every step from one hop to the other started to shorten until it reduces dramatically while the letter is closer to the target. This because apparently at the very beginning of this process we are interested to send the letter the farthest away, but while we approach the target we need to apply a more precise strategy to not send the letter far away, **so there is a geographical link to this idea** and we want to exploit this strategy.

99.3 An Algorithmic Perspective (GeoGreedy Algorithm)

The GeoGreedy algorithm works in such steps:

- Starting node s must send a message to target node t
- Location of t in the grid is known to s
- s has no knowledge of any intermediate edges, **but its local contacts and its long range contact**, so we assume that we have a lot of local contacts and at least one contact that covers a long range.
- every **intermediate node will try to route the message as closer as possible to t**

The problem here is that weak ties (random edges) are too random in the Watts Strogartz model because they are created from one point to another totally random point in the network, and this is why it fails to represent efficiently the decentralized search. We want now to assess this randomness in a way that is more bounded to the distance and to do this we will use the generalized network model.

99.4 Navigation in a Small World

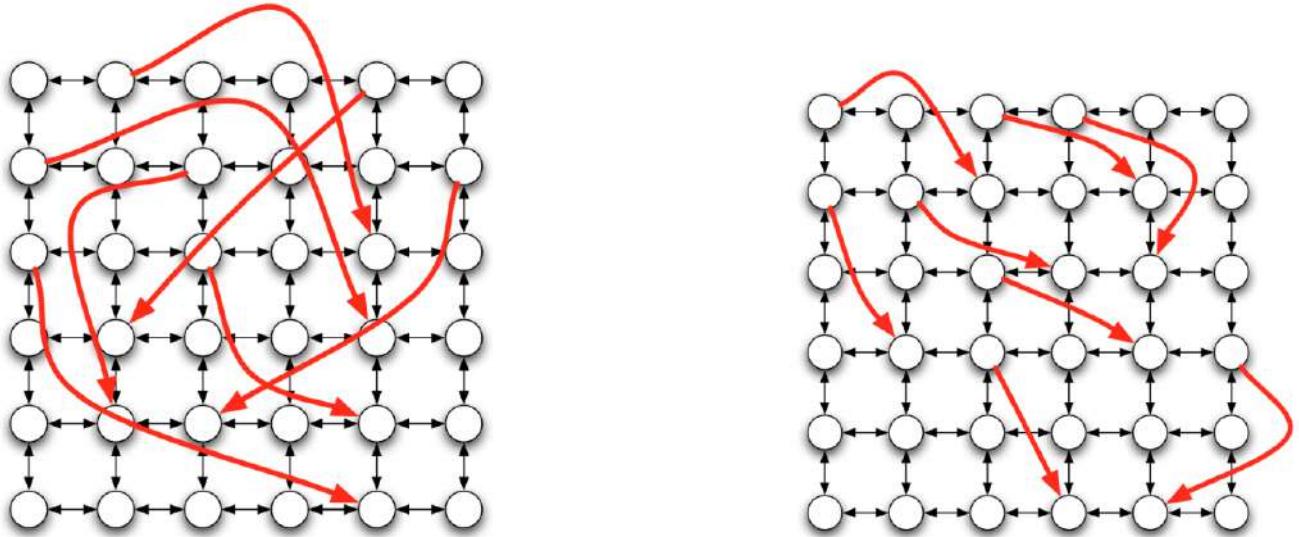
We introduce a **clustering exponent q** and use it in this steps:

- The random edges are generated in a way that decays with distance (d)
- Given v and w , with $d(v,w)$ we denote their distance if one has to walk along adjacent nodes on the grid and this distance is calculated according a path that one has to walk among adjacent nodes on the grid
- Randomly generated edge **out of v links to w with probability proportional to $d(v,w)^{-q}$** that wasn't present in the original Watts Strogartz model. In other words this means that the distance is linked to w with a probability that is inversely proportional to the distance up to an exponent q .

In the original Watts Strogartz model this exponent q was 0 since links are chosen uniformly at random.

99.5 Comparing Models with Different Clustering Exponents

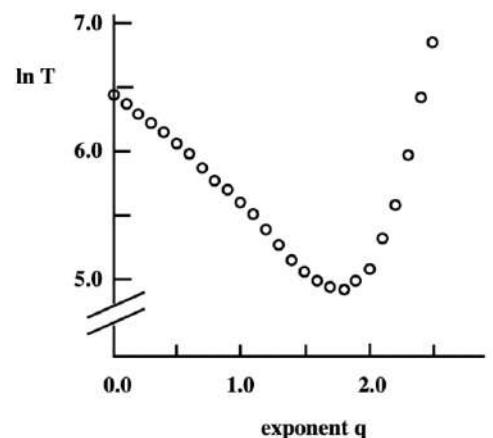
In original Watts Strogartz model nodes can connect very distant links since endpoints are chosen randomly and the clustering exponent q is zero. While in the GeoGreedy algorithm and according to a small exponent q and we have that the probability to link two nodes that are not that distance is higher compared to the probability of linking two very distant nodes.



This is quite a natural thing, since we have a much higher chance to link with a person that lives closer to use instead of someone that lives in New York.

99.6 Estimate of the Clustering Exponent q by Simulation

- ◊ simulations of 1000 runs of decentralized search
- ◊ a (kind of) grid with 400 million nodes
- ◊ T : delivery time, i.e., the averaged number of steps required to reach the target
- ◊ $q=0$ (WS model): T is short, but far from being "the shortest"
- ◊ best results for $1.5 \leq q \leq 2$
- ◊ As the network size increases, best performances when $q \rightarrow 2$
- ◊ \Rightarrow the random links follow a inverse-square distribution d^{-2}



From: Easley, Kleinberg [ns2]

We found a way to optimize this value of q with the GeoGreedy decentralized search algorithm so we have a qualitative and quantitative definition of this procedure.

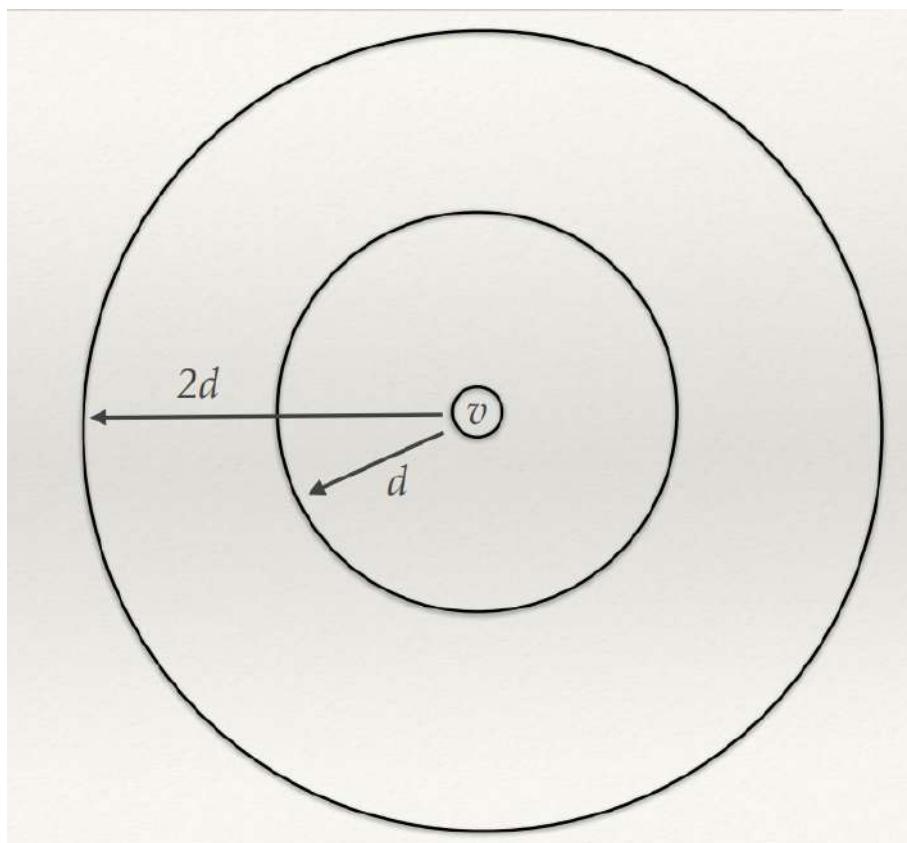
99.7 Intuitive Explanation of $q = 2$

If we are in a context of large networks we can agree that we organize distances into different scales of resolution like world, state, town etc.. From the perspective of node v we have that groups of nodes at increasingly large ranges of distances from v can be also groups at different scales of resolution. The idea is that Milgram's letters reduced their distances to the target by approx. a factor of 2 at each step, so the distances were just divided by an half while the letter was approaching the target.

If we consider **concentric scales of resolution around v** , we can study how the interacts with scales of resolution from node v perspective.

If we consider nodes at distances between d and $2d$ from v we have somewhat of a circle, where we have a radius and a circumference and the area of the circle inside contains all the people that are connected to v at distance d , or $2d$ if we consider a larger circle. We can then make then some statements:

- the **number of nodes** is proportional to d^2 (area square of the radius)
- the **probability of linking to each node in the group** is proportional to $1/d^2$ which is d^{-2}
- the probability that a random edge links into some node in this ring is approximately independent of the value of d , because if we want to get someone in particular it is not really dependent of the distance, but if we want to find every one else in the outer link, the probability is inversely proportional to the square of the distance.



99.8 Qualitative Interpretation of $q = 2$

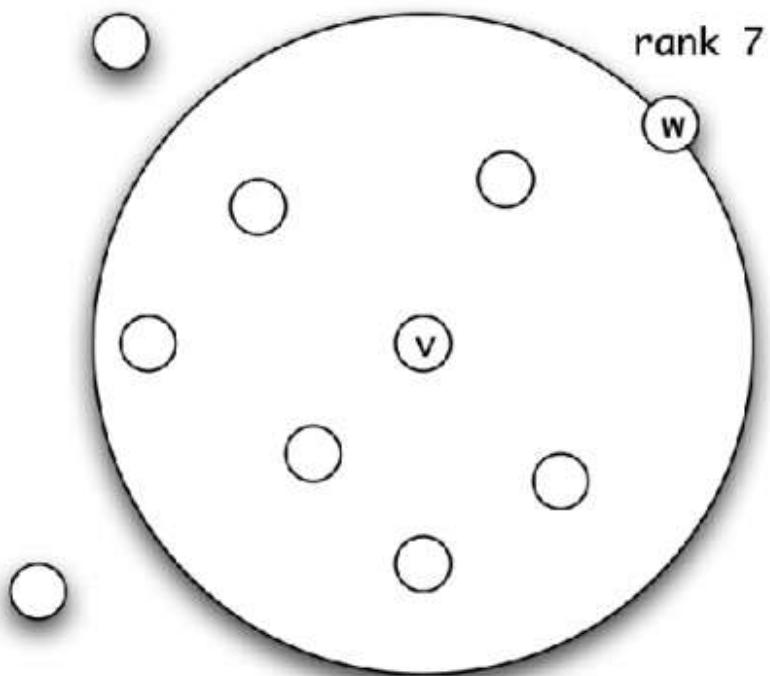
Long range contacts formed in a roughly uniform way when $q = 2$. People forward the message reducing consistently the distance to the target, no matter how near or how far they are from it. However, the postal system has been designed centrally and the pattern through the inverse-square network arise spontaneously from a completely random pattern of links.

100 EMPIRICAL ANALYSIS AND GENERALIZED MODELS

A group of researches started digging into a blogging site and the analysis was done on 500,000 users, with geographical info, and links to their friends in this blog. The structure of friendships in online communities was used to understand the links even in the real offline world. The problem of this was that the population density was non-uniform, since even the distribution of people in the US is not uniform since there are much more people on the east side of the states than on the west side. And for this reason we cannot use the geographical distance as a reliable measure, and the idea of this research was to use a rank instead of this distance.

100.1 Rank-Based Friendships

As said before the idea was to determine link probabilities by rank and not physical distance. The v node's perspective is that the $\text{rank}(w)$ of another node w is the number of nodes that are closer to v than w is. **This is rank-based friendship with exponent p ,** where v creates a random link to a chosen node w with probability proportional to $\text{rank}(w)^{-p}$. It is just a way to rewrite the generalized model not using distances but ranks.

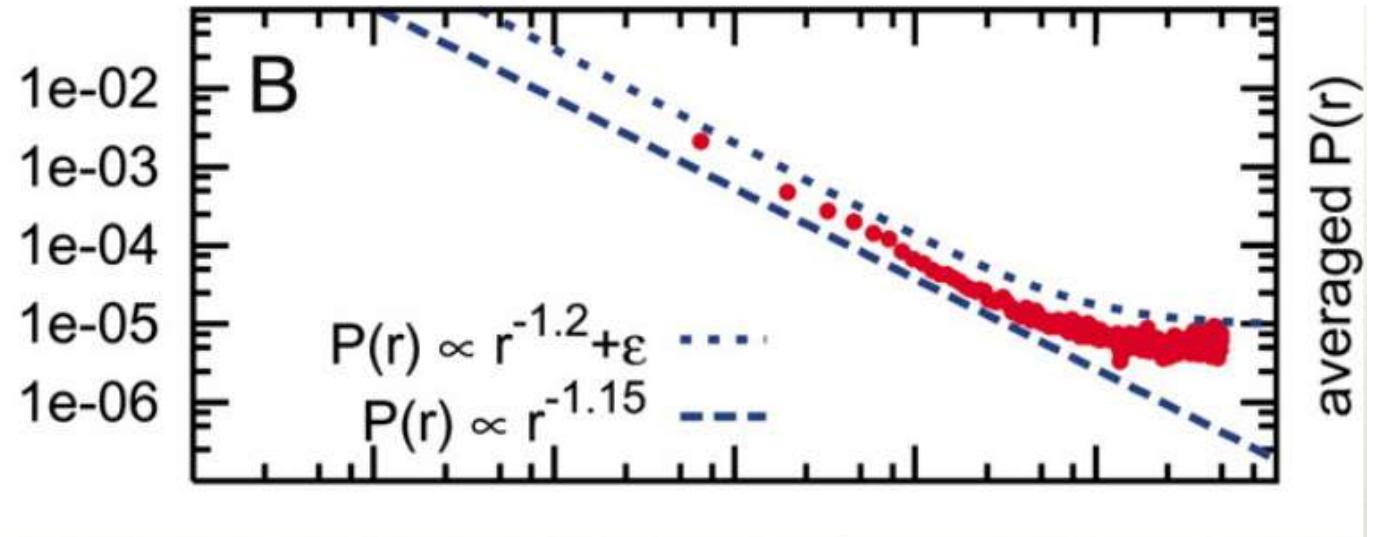


(a) w is the 7th closest node to v .

100.2 Ranks and GeoGreedy

We want to understand which choice of p to generalize the inverse-square distribution for uniformly spaced nodes.

If w is at distance d from v , than it is on the circumference of radius d , that contains d^2 closer nodes, and the rank is $\text{rank}(w) = d^2$, and this means that linking with w has probability proportional to d^{-2} and equal to $\text{rank}(w)^{-1}$, since we have to subtract one from the exponent and hence $p = 1$ could be the right generalization.



The **empirical validation** is given by understanding the fraction of a friendship as a function of geographic rank on the blog. The data is between two straight lines with slopes -1.15 and -1.2 (very close to -1), and according to the fact that the rank up to -1 can be approximated to the inverse of the square of the distance d^{-2} we have that this blog is a rank based network and it is searchable through this algorithmic perspective.

100.3 Social Foci and Social Distance

Social Foci says that any type of communities, such as groups, neighborhood, shared interest, or activity it is likely that the shared foci with only a few members are the strongest generators of new social ties. In other words, if the community is small the links inside of this community will be a lot and the community itself will have an higher density.

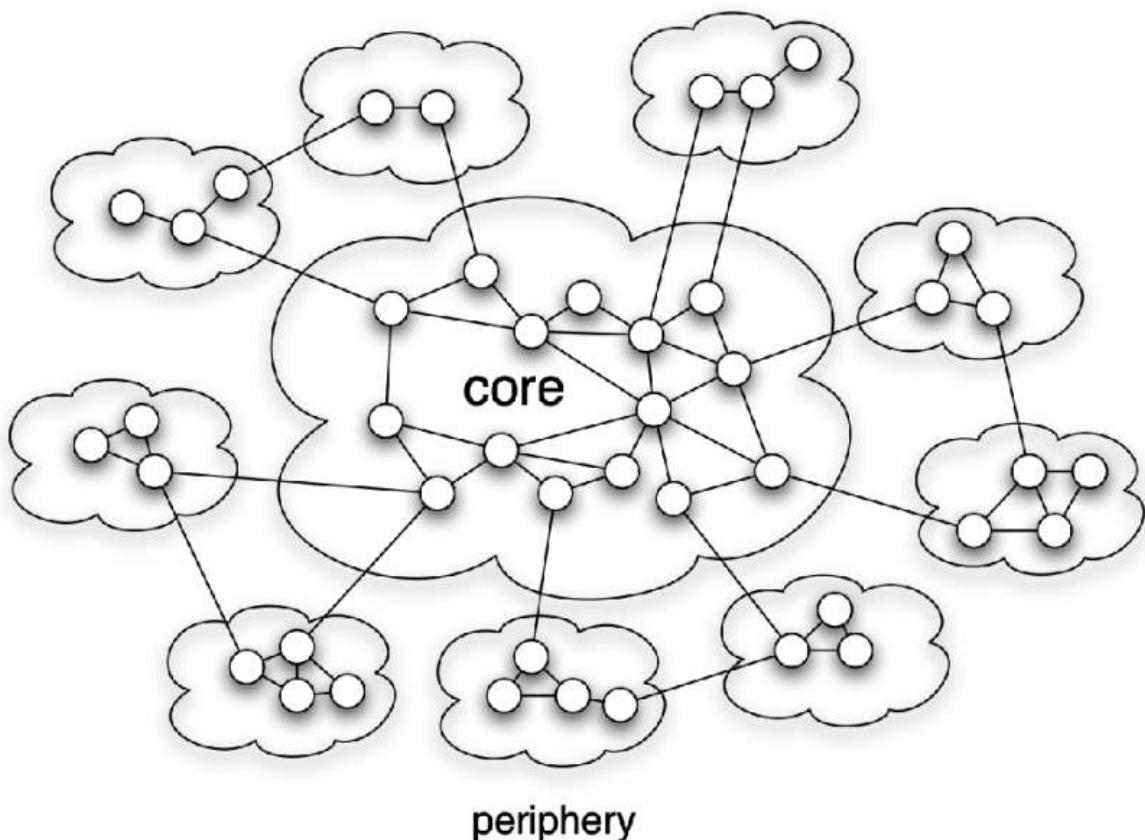
A natural definition of **social distance** $\text{dist}(v,w)$ between two people is **the size of the smallest focus** that includes both of them. If we find another person that belongs to a social circle to which both of us belong, the size of that social circle is the distance between me and the other person and in order to understand the shortest path to be introduced to this person, we should try and understand which is the smaller social circle to which both of us belong. And now we create a link between v and w with probability proportional to the $\text{dist}(v,w)^{-p}$.

If we can prove that $p = 1$ we have as a result that the network is efficiently searchable, for example if we want to know some one that is inside our smaller social circle the task is quite easy, while if the circle is bigger, the task will be harder.

101 CORE PERIPHERY STRUCTURES AND DIFFICULTIES IN DECENTRALIZED SEARCH

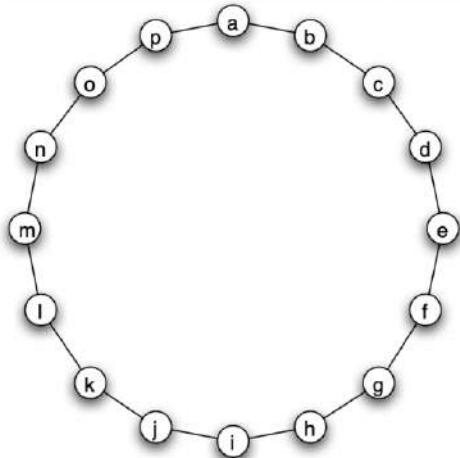
We know that decentralized searches is most successful when the target person is well known and has a socially high status, this is natural and even if finding a short path to Berlusconi looks counter intuitive, it is much easier then finding a shortest path to a person that no one knows.

Homophily suggests assortativity by status which means that large social networks tend to be organized in core-periphery structures, so the famous people are connected to other famous people and at the periphery we can find less famous people connected to even lesser known people. High-status people are linked in a densely connected core, and also establish links that span geographic and social boundaries, while low-status people tend to form links that are much more clustered and local, hence reaching someone in the core is much easier then reaching someone in the periphery.

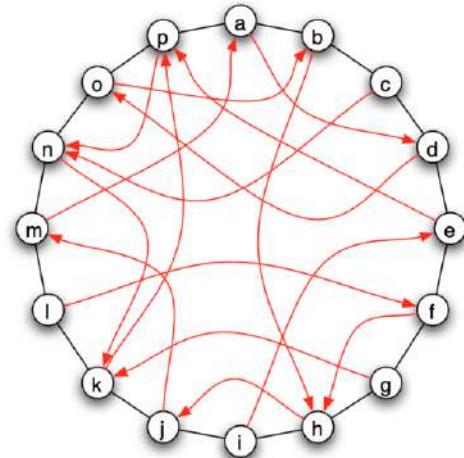


102 ANALYSIS OF DECENTRALIZED SEARCH

We want to understand formally why the search is successful with d^{-2} distributions. We want to prove that in a one dimensional ring the best exponent $q = 1$, and once this is proved we can shift the argument for dimensions equal or greater than 2 claiming that the best exponent is $q = 2$, and at the end we will prove why this choice of q is the best for decentralized search in the limit of increasing network size.



(a) A set of nodes arranged in a ring.



(b) A ring augmented with random long-range links.

We start with a one dimensional ring that has its neighbors connected but it has a long range connection with a random node in the ring. If we want to reach a very distant node, we can start following the long range link and then understand how we can get closer to our target following the current node perspective.

We need to formalize a bit better the topology of the ring. We will adopt the **Myopic Search** that is the GeoGreedy algorithm but in just one dimension. We have a set of n nodes in a ring, where each node v is connected by directed edges to:

- the **two local contacts**, which are the neighbors
- to a **long-range contact**, and the probability that connects to a given w is proportional to the inverse of the distance $d(v, w)^{-1}$

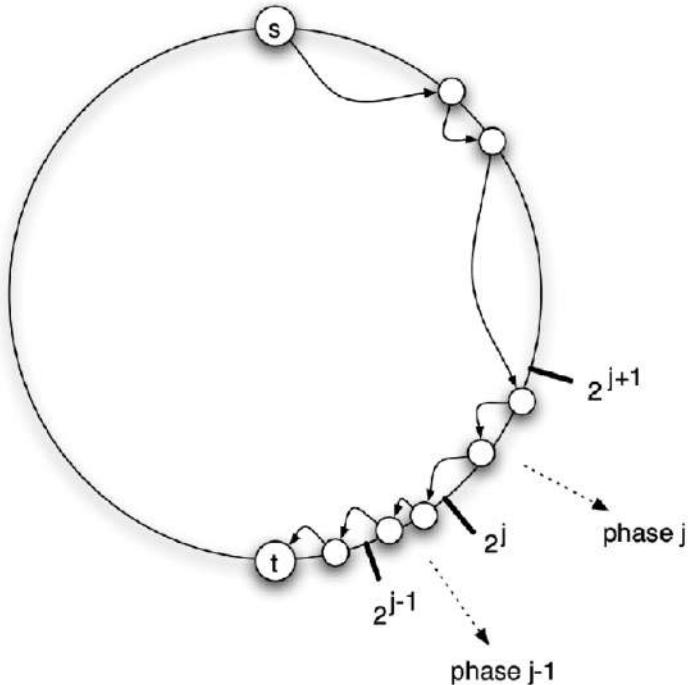
The Myopic search is done by selecting a starting node a and a target node i , but we need to precise that this **myopic search does not lead to the shortest path from a to i** , but we can still expect short paths in general.

CONTINUES TO THE NEXT PAGE

The steps we need to follow are:

- We generate a random network by adding long-range edges to a ring
- We choose a start node s and random target node t
- The number of required steps is just a random variable X
- We want to understand the expected value of the random number of steps X is $E[X]$
- At the end we want to prove that this number of steps $E[X]$ is relatively small

We start by tracking how long it takes for the message to reduce its distance by factors of 2 as it closes to the target. If the message is in **phase j** of the search if its distance from the target is between 2^j and 2^{j+1} . The total number of different phases is $\log_2 n$.



- We can write the search total time as the sum of times taken in each phase

$$X = X_1 + X_2 + \dots + X_{\log_2 n}$$
- Linearity of expectations \Rightarrow the expectation of a sum of random variables is equal to the sum of their individual expectations:

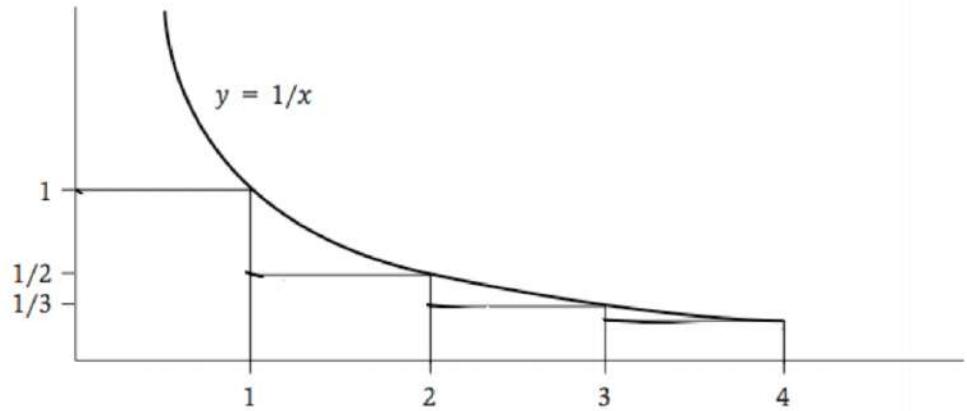
$$E[X] = E[X_1 + X_2 + \dots + X_{\log_2 n}] = E[X_1] + E[X_2] + \dots + E[X_{\log_2 n}]$$
- We want to show that $\forall j : E[X_j] \propto \log_2 n$
- We will prove that $E[X] \propto (\log_2 n)^2$, i.e., the myopic search constructs a path that is *exponentially smaller* of the number of nodes

We need the **Normalization Constant**

- ◊ we are repeating that v forms its long-range link to w with probability *proportional to* $d(v, w)^{-1}$
 - ◊ we know the set of probabilities: we can set the missing constant of proportionality to $1/Z$, such that Z is a normalizing constant
- $$Z = \sum_{u,v:u \neq v} d(u, v)^{-1}$$
- ◊ Notice that we have: 2 nodes at distance 1 from v , 2 nodes at distance 2, ..., 2 nodes at distance $n/2$:

$$Z \leq 2 \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{n/2} \right)$$

- ◊ We have the sum of k reciprocals (for $k = n/2$), whose upper bound is the curve $y = 1/x$



$$\diamond 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{n/2} \leq 1 + \int_1^{n/2} \frac{1}{x} dx = 1 + \ln(n/2)$$

- Let's go back to Z estimate:

$$Z \leq 2(1 + \ln(n/2)) = 2 + 2 \cdot \ln(n/2)$$

- In order to simplify our bound, observing that $\ln x \leq \log_2 x$:

$$Z \leq 2 + 2 \cdot \log_2(n/2) = 2 + 2(\log_2 n) - 2(\log_2 2) = 2(\log_2 n)$$

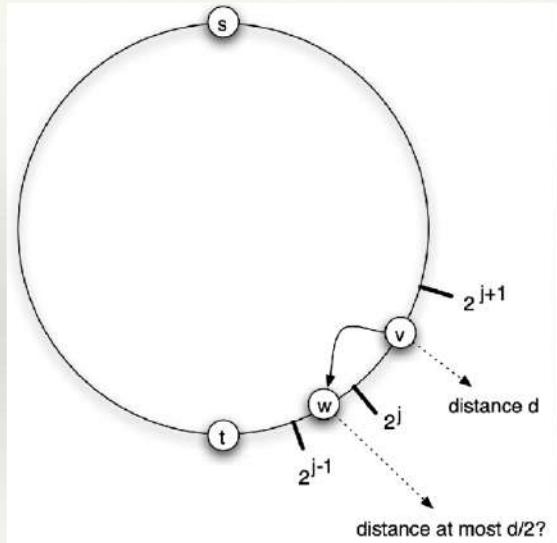
- Finally, we have that the **actual probability** that v links to w is:

$$\frac{1}{Z} d(v, w)^{-1} \geq \frac{1}{2\log_2 n} d(v, w)^{-1}$$

(*) Observe: here we used the fact of being in one dimension

102.1 Time Spent in One Phase

- Now, we need to show that the time spent by the search in any phase j is small
 - message at v is at a distance d between 2^j and 2^{j+1} to t
 - this phase will be ended when the distance to t decreases below 2^j
- If after that the message has taken v 's long-range contact, we observe that distance to target will be reduced from d to $\leq d/2$, then v will be the last node to belong to phase j .
 - is the probability for this to happen small enough?



- Let I be the set of nodes at distance $\leq d/2$ to t
- number of nodes in I : $d+1$, including node t itself, and $d/2$ nodes consecutively on each side of it
- Each node w in I
 - has distance at most $3d/2$ from v
 - has probability of being the long range contact of v at least:

$$\frac{1}{2\log_2 n} d(v, w)^{-1} \geq \frac{1}{2\log_2 n} \cdot \frac{1}{3d/2} = \frac{1}{3d \cdot \log_2 n}$$

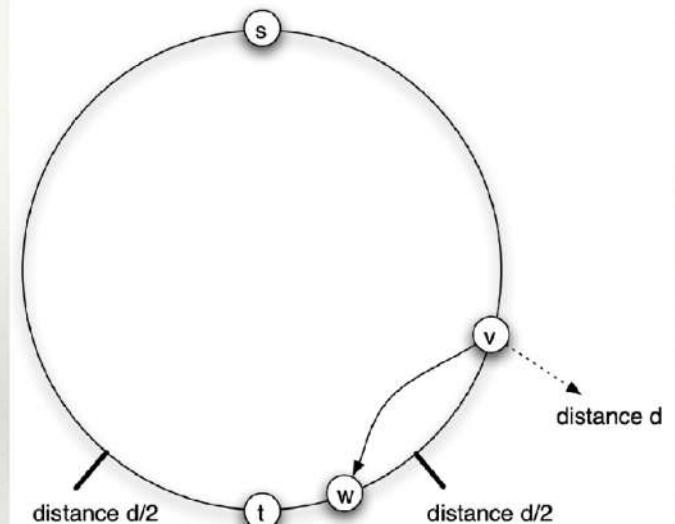


Figure from: Easley, Kleinberg [ns2]

- There are at most d nodes in I , then the probability that *one of them* is the long-range contact of v is at least:
- $$d \cdot \frac{1}{3d \cdot \log_2 n} = \frac{1}{3\log_2 n}$$
- (**) Observe: here we used again the fact of being in one dimension
- Notice that this is proportional to $\log_2 n$ **regardless of the value of d**
 - If one of these nodes is the long-range contact of v , than phase j ends
 - otherwise, to run for at least i steps, the process has to fail to exit this phase a number of $i-1$ times; prob. that phase j runs for at least i steps is at most:

$$\left(1 - \frac{1}{3\log_2 n}\right)^{i-1}$$

- Let's use the formula for the expected value of a random variable:

$$E[X_j] = 1 \cdot P(X_j = 1) + 2 \cdot P(X_j = 2) + 3 \cdot P(X_j = 3) + \dots$$
- Notice that this is equivalent to:

$$E[X_j] = P(X_1 \geq 1) + P(X_2 \geq 2) + P(X_3 \geq 3) + \dots$$
- Since we have that:

$$P(X_j \geq i) \leq \left(1 - \frac{1}{3\log_2 n}\right)^{i-1}$$

- Then

$$E[X_j] \leq 1 + \left(1 - \frac{1}{3\log_2 n}\right) + \left(1 - \frac{1}{3\log_2 n}\right)^2 + \left(1 - \frac{1}{3\log_2 n}\right)^3 + \dots$$

- The geometric sum with multiplier $\left(1 - \frac{1}{3\log_2 n}\right)$ converges to

$$\frac{1}{1 - \left(1 - \frac{1}{3\log_2 n}\right)} = 3\log_2 n$$
- We are almost there: we have that $E[X_j] \leq 3\log_2 n$
- and finally:

$$E[X] = \sum_j^{log_2 n} E[X_j] \leq 3(\log_2 n)^2$$

a quantity proportional to $(\log_2 n)^2$ as we wanted to show

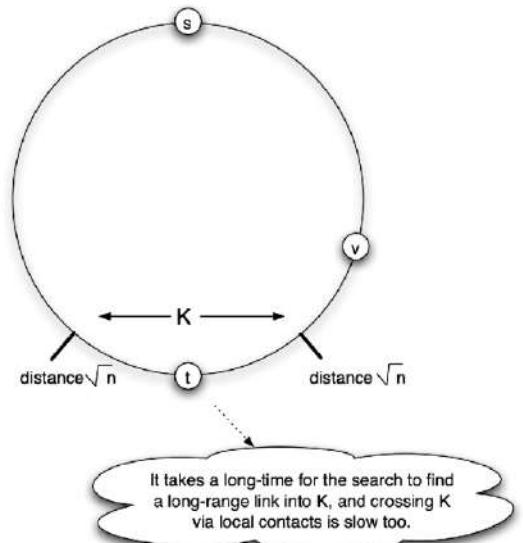
102.2 Analysis in Two Dimensions

- ◊ We used the fact that we were in one dimension in (*) and (**).
 - ◊ (*) to calculate constant Z
 - ◊ (**) to argue that there were at least d nodes within distance $d/2$ of the target t
- ◊ We can repeat the calculation of Z to find that it still proportional to $\log_2 n$ when the prob of v linking to w is proportional to $d(v, w)^{-2}$ (i.e., $q = 2$)
- ◊ We can argue that the number of nodes within distance $d/2$ is proportional to d^2
- ◊ Therefore, the prob. that the message halves its distance to the target in this step is at least $d^2/(d^2 \log_2 n) = 1/(\log_2 n)$: the rest of the analysis is the same
- ◊ We can also argument that networks built by adding long-range contacts to grids in $D > 2$ dimensions are searchable when the exponent $q = D$

Decentralized Search is Less Efficient for $q \neq 2$

- ◊ with $q = 0$ (WS model)
- ◊ Starting node is outside K
- ◊ Long-range edges are created uniformly at random, the prob. that any one node has a long-range contact inside K is equal to the size of K divided by n :

$$\leq 2\sqrt{n}/n = 2/\sqrt{n}$$
- ◊ Therefore, search will expected to take $\geq \sqrt{n}/2$ steps to find a node with a long-range link leading into K
- ◊ Hence, it is hard for the search to enter K
- ◊ Argument for too much high exponents: long-range links are too short, and search takes a long time to find links that span sufficiently long distances



103 EPIDEMICS - DISEASES AND THE NETWORKS THAT TRANSMIT THEM

We know that important diseases have characterized human history from ancient times. We have a certain estimation of the deaths that these diseases have caused such as some historical plagues that have caused up to 50 millions of people, that was 40 percent of the population at that time in the 1500.

If we want to characterize ancient diseases we can observe that they started in some crucial areas in the south of Europe, and then they kept spreading up towards the north for 5 or more years. This differentiates ancient diseases from modern ones since we can estimate the speed of contagion and spreading and we can define this speed as the speed of a walking human, since it is the main transporter of the viruses, and the humans can bring the disease in geographical points that are very far apart.

103.1 Contact Networks

The quickness of the spreading of the virus is due to the contact networks that we have in real life, since if we are close to somebody in real life it is quite possible that we either catch or transmit a virus. The basic idea that we want to explore is that trying to **accurately modeling the underlying network is crucial to understanding the spread of an epidemic**, in particular in the early stages of the epidemic itself, when the contagions and speed of spreading can be trackable.

Another thing that characterizes modern epidemics from ancient ones is that the connections between very distant places in the world are everywhere since we have **airports, highways, ferries, boats** and all the means to travel from a distant point to another. This possibility of moving around the world of course impacts the quickness of the spreading process of an epidemic and needs to be considered when we want to track down its process.

103.2 Simple vs Complex Contagion Models

Ideas, behaviors and infectious diseases all spread in very different ways, and we can classify contagion between **simple and complex**:

- both types spread from person to person
- with **social contagion** people make decision to adopt a new idea or an innovation
- with **biological contagion** it is not dependent on the individual's will, and it is not easily observable at the person-to-person interaction level. If a little patogen is going to be transmitted from A to B, this transmission cannot be detected while it is happening.

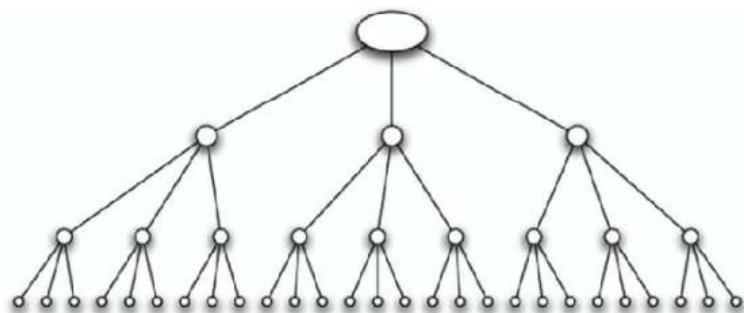
Infection diseases can be seen as a simple contagion processes; however, due to not trivial observability, it is usually better to adopt **non deterministic models which are based on randomness and probabilities**. It is different compared to the social contagious models since we need to add probabilities in our models and compare them to baselines where hypothesis based on randomness are made.

104 BRANCHING PROCESS

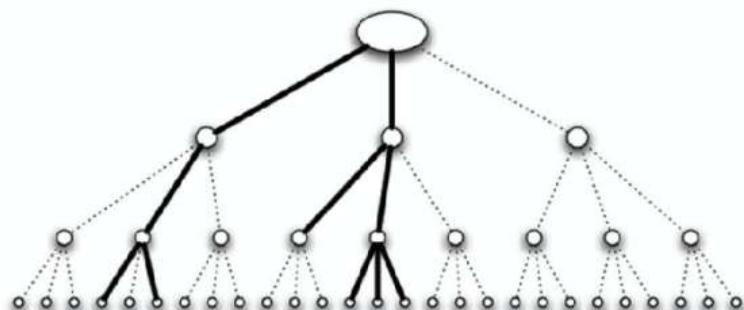
The **branching process** is the simplest form of contagion and the assumption is that the underlying network, in which we are studying the spread of the virus, is **an infinite tree**, that of course is only a theoretical assumption. We need to make some other assumptions to work with this model:

- **The Patient 0**, who is the first person that contracted the virus, is the **root** of the tree
- A node at **level n** (representing a wave of the epidemic) is a person that meets other **k** people
- Every person passes the disease with **probability β** .

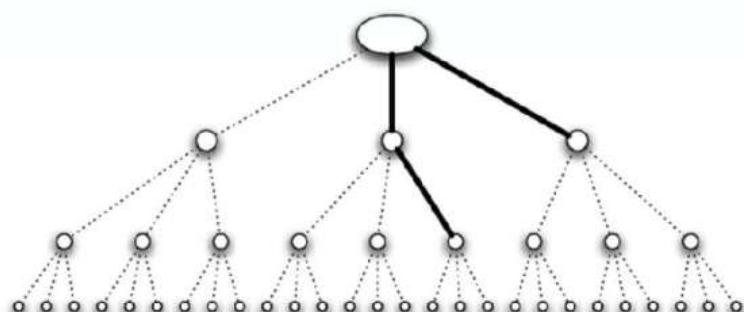
There are many other factors to consider such as the immune system of the people and etc.. to keep things simple for this model we assume that every person can pass the disease to its contacts in the contact network. With all these assumptions we want to understand if we can predict if and when the disease vanishes. If we can do it in a simple model such as this one, we can extend this model to much more complicated ones.



(a) *The contact network for a branching process*



(b) *With high contagion probability, the infection spreads widely*



(c) *With low contagion probability, the infection is likely to die out quickly*

104.1 The Basic Reproductive Number r_0

The idea behind this number, that is also a main characteristic in a branching process, is that if all the individuals in wave n fail to infect any other individuals, the infection vanishes from waves $\geq (n + 1)$. We can try to understand and compute the probability that this is going to happen at a certain wave using R_0 .

R_0 is the basic reproductive number of the disease is the expected number of new cases of the disease caused by a single individual, and in a branching process this number is equal to βk , that is the transmission probability multiplied by the k people we will be in contact with.

We want to prove that:

- if $R_0 < 1$ then with probability 1, the disease dies out after a finite number of waves, and hence we can not worry about this virus since it won't spread to become an epidemic
- if $R_0 > 1$ then with probability greater than 0 the disease persists by infecting at least one person in each wave, in other words if R_0 is greater than 1 there is a not neglectable possibility that the virus will spread globally

R_0 can be applied as a rule of the thumb, but with some precautions:

- It can be estimated empirically, and it is a useful approximate indication of the spreading power of the disease, when epidemiologists do not have other precise conditions governing the epidemic
- The tree is far from being representative of a real contact network, so these results are applied to ideal networks

When R_0 is close to 1, and the contagion probability increases or decreases of a little bit, the epidemics can have the opportunity to become an enormous outbreak or to consistently reduce its danger and spreadness. If we consider branching networks and remember that $R_0 = \beta k$ we know that small changes in k can have a large effect on R_0 . If we follow the idea of the butterfly effect, we know that even the smallest of changes in the number of contacts, can end up having a massive impact in the whole world.

The basic reproductive number (R_0), and the real-life effective reproductive number (R_t) for a given population are different:

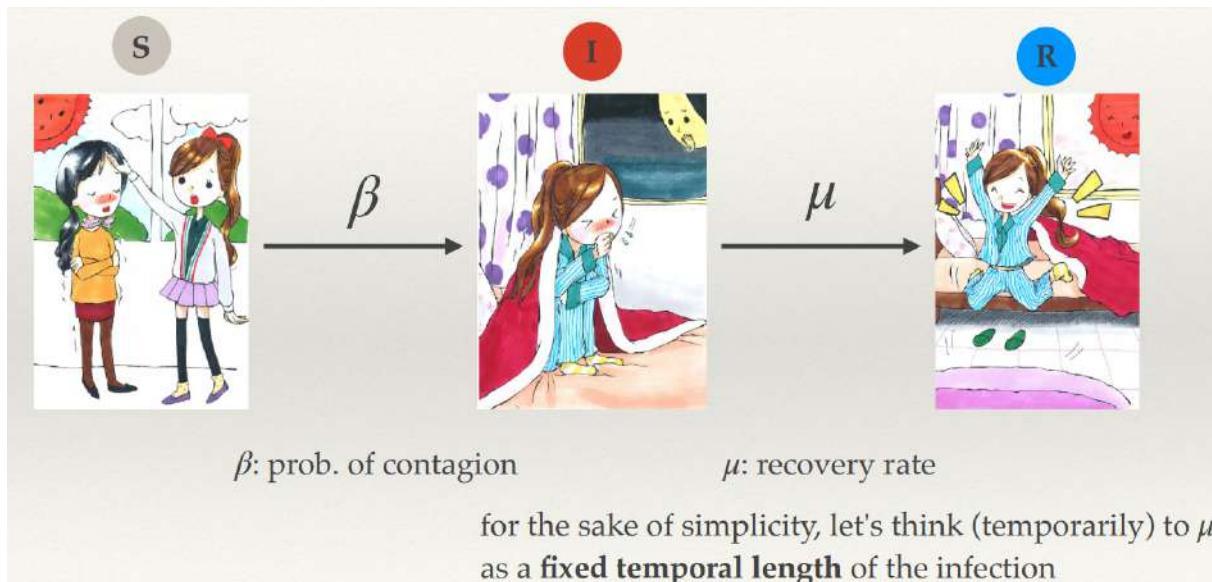
- R_0 is the number of secondary cases generated by the presence of one infected individual in an otherwise fully susceptible, well-mixed population.
- R_t is a more practical real-life version of this, which uses real-life data (from diagnostic testing and/or clinical surveillance, vaccines etc..) to estimate the reproductive number for an ongoing epidemic.

From a biological point of view R_0 is important, but we should not use R_0 to make a strategy since it does not take into considerations the policies and restrictions that the governments adopt, but it is a warning that indicates that if we don't do anything to cure a certain virus, then it will keep spreading.

105 THE SIR EPIDEMIC MODEL

It is a way to pass from trees to networks to have a more realistic scenario. Branching processes are a good starting point when we do not know anything about the network structure, but we need models that can be applied to all the types of networks. We need to introduce some states for this model:

- **Susceptible:** a node with such state has not caught the disease yet
- **Infectious:** a node with such state has caught the disease and has some probability of infecting others
- **Removed:** after an infectious period, the node has recovered, and it will no longer infect anyone else or be infected again, maybe even because it died.



105.1 SIR Infections in Networks

In order to use the SIR model in a network we need to make some approximational assumptions such as:

- The underlying contact network is a **directed graph** that considers link that have a direction
- We can setup an **Initialization process** such that some nodes are in the I state and all the others on the S state
- The infection itself can be simulated saying that each node that enters the I state remains infected for a fixed number of steps μ (μ), that can be a set number of days, and it passes the disease to each of its susceptible neighbors with probability beta.
- We can simulate **Recovering** saying that after μ steps, the infected node is no longer infectious or susceptible, and we describe it as **removed (R)**, since it is now an inert node in the contact network. The R nodes are very useful since they can act as a barrier that protects other nodes from being infected.

If $\mu = 1$, at time $t = 0$ we can have a lot of susceptible nodes and a couple of infected nodes. At time $t + 1$ the previously infected nodes are recovered and with a certain probability some other susceptible nodes contracted the diseases and are now infected. With $t = 2$ the process of infection goes on and we can arrive at a certain time t when the infected nodes become Removed and the diseases vanished from the network.

We can make some observations now:

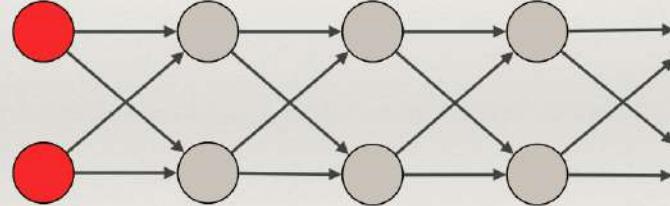
- ◊ The branching process is a **special case** of the SIR model, where:
 - ◊ the contact network is an infinite tree, with each node connected to a fixed number of child nodes
 - ◊ and $\mu = 1$
- ◊ To study the general dynamics:
 - ◊ execute different **agent-based simulations** (each simulation is a realization of the model)
 - ◊ observe what happens to $\frac{S(t)}{N}$, $\frac{I(t)}{N}$ and $\frac{R(t)}{N}$ when $t \rightarrow \infty$
 - ◊ otherwise, solve the problem mathematically (next lecture)

105.2 The Role of r_0

Some nodes can be protected inside of their own clusters compared to other clusters where the disease is spreading quicker. **Weak ties** in this situation can have a very important role since if we are infected by a weak tie, we have an high chance of bringing the disease inside of our cluster. The initial structure can have a massive impact on the final state of the network.

- With an arbitrary network structure, the dichotomy in epidemic behavior determined by R_0 does not necessarily hold (\rightarrow next lecture)

- example: $\beta = 2/3, \mu = 1$



- $R_0 = 2 \cdot \frac{2}{3} = \frac{4}{3} > 1$; nevertheless, the disease will vanish from the above network.

each edge fails to transmit with prob. $(1 - \beta) = \frac{1}{3}$

All four edges fail to transmit with prob. $\left(\frac{1}{3}\right)^4 = \frac{1}{81}$

Sooner or later this will happen in a **number of finite steps**: with prob. 1 the disease will die out!

This means that there will be a wave where the disease will die out. The consequence is that we calculated R_0 and found out that it was greater than 1, and hence we can expect that the disease won't die out immediately, but we know that at least with a given network topology, the disease will die out.

105.3 SIR Epidemics and Percolation

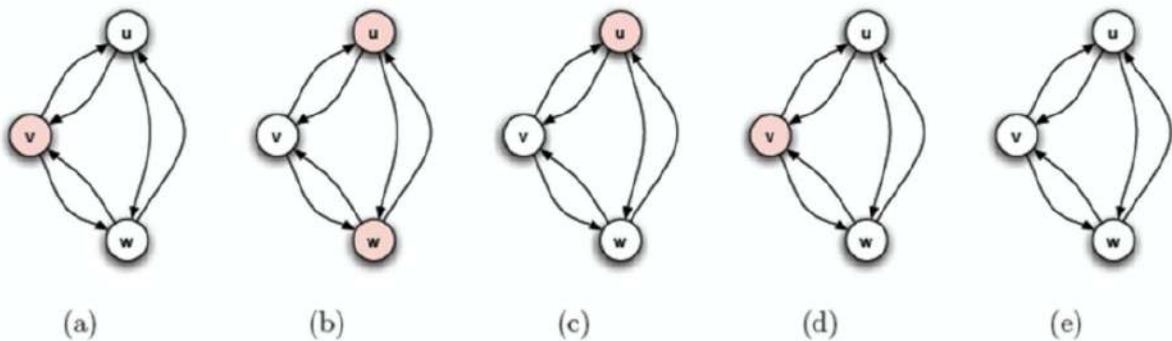
We can give a static view of the process, again focusing on the SIR model in which $\mu = 1$:

Instead of deciding with probability beta at time t if the edge will transmit the disease, we can run our probabilities at the beginning and this will **open or block** the edges then, we can just check if there are paths from the infected nodes to other nodes in the network to understand which nodes will be infected and this it is like observing fluid percolation in a system of pipes.

106 THE SIS EPIDEMIC MODEL

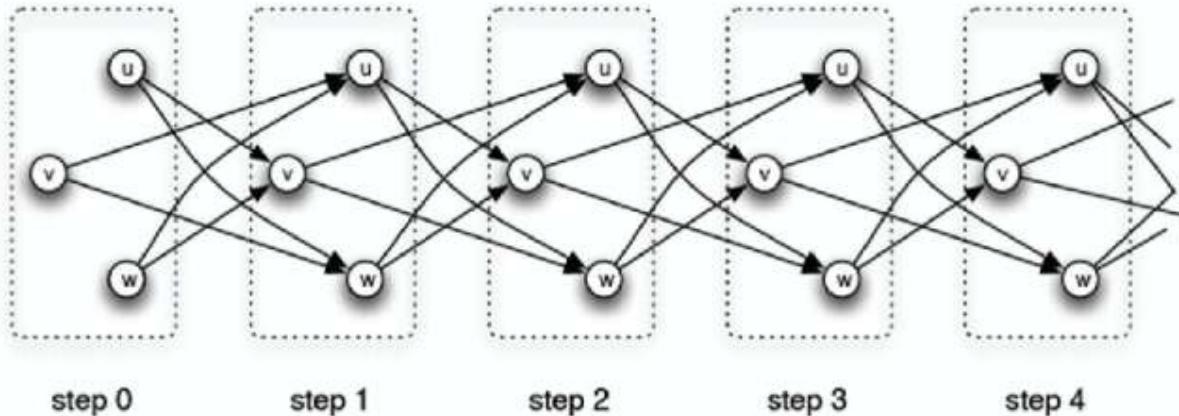
In this model we introduce the assumption that **an infected node can go back to a susceptible state**. This means that we can actually be infected again after we have recovered, and hence in this model there isn't a recovered state as in the SIR model.

- In this model we only have **susceptible and infectious states**
- **Initialization** is with some nodes are in the I state and all the others on the S state, as in the SIR model
- **Infection** happens when each node that enters the I state remains infected for a fixed number of steps μ , and it passes the disease to each of its susceptible neighbors with probability β .
- **Recovering** is modeled saying that after μ steps, the infected node is no longer infectious returns to the S state

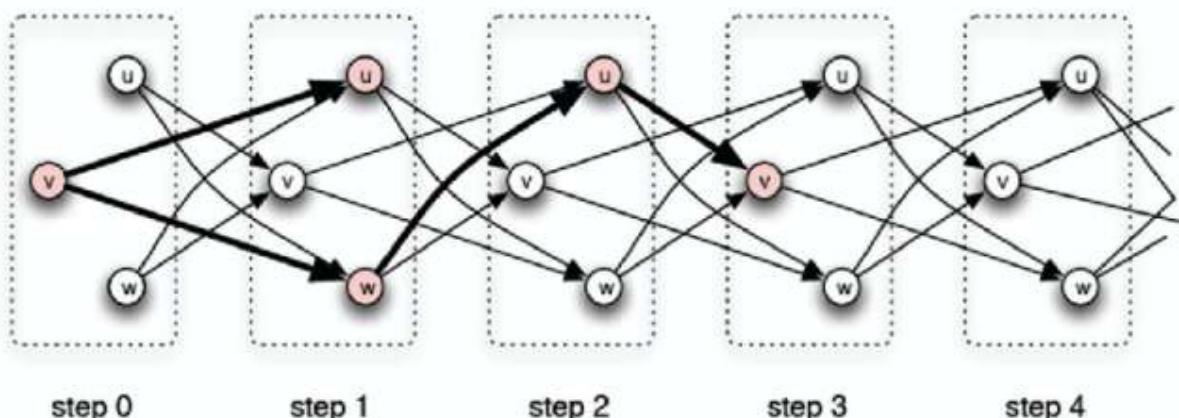


106.1 Connections Between SIR and SIS

There are connections between the two different models and we can use a multi layer approach to see that we can represent the SIS model with the SIR one, where each layer of the network represents a different timestamp of the contact network.



(a) To represent the SIS epidemic using the SIR model, we use a “time-expanded” contact network



(b) The SIS epidemic can then be represented as an SIR epidemic on this time-expanded network.

Figure 21.6: An SIS epidemic can be represented in the SIR model by creating a separate copy of the contact network for each time step: a node at time t can infect its contact neighbors at time $t + 1$.

107 SYNCHRONIZATION AND OSCILLATIONS

Synchronization and oscillation are quite important in diseases since they could be linked to seasonality, if we think about the flu it is much more common in the winter and a lot of states show the same dynamics of this epidemic even if the states are far apart.

In order to study the global dynamics of a disease, it is worth **studying the tendency of epidemics for certain diseases to synchronize across a population, sometimes producing strong oscillations in the number of affected individuals over time.**

We observe seasonality and periodical oscillations but still today we don't have an explanation for this phenomena, some hypothesis that exploit networks can be made though:

- they could happen due to large scale societal changes or
- contagion dynamics of the disease itself

We can try to model this by trying to embed a notion of **temporary immunity and long-range links in the contact network.**

108 THE SIRS EPIDEMIC MODEL

We can combine elements of the SIR and SIS models to confer temporary immunity on infected individuals, which means:

- The **Initialization** is the same, where some nodes are in the I state and all the others on the S state
- But the **Infection dynamic changes**, since each node that enters the I state remains infected for a fixed number of steps μ_i , and it passes the disease to each of its susceptible neighbors with probability beta.
- The **new feature** in this model is that after μ_i steps the infected node is no longer infectious. It enters the Recovered state for a fixed number of steps μ_r , then it returns to the S state.

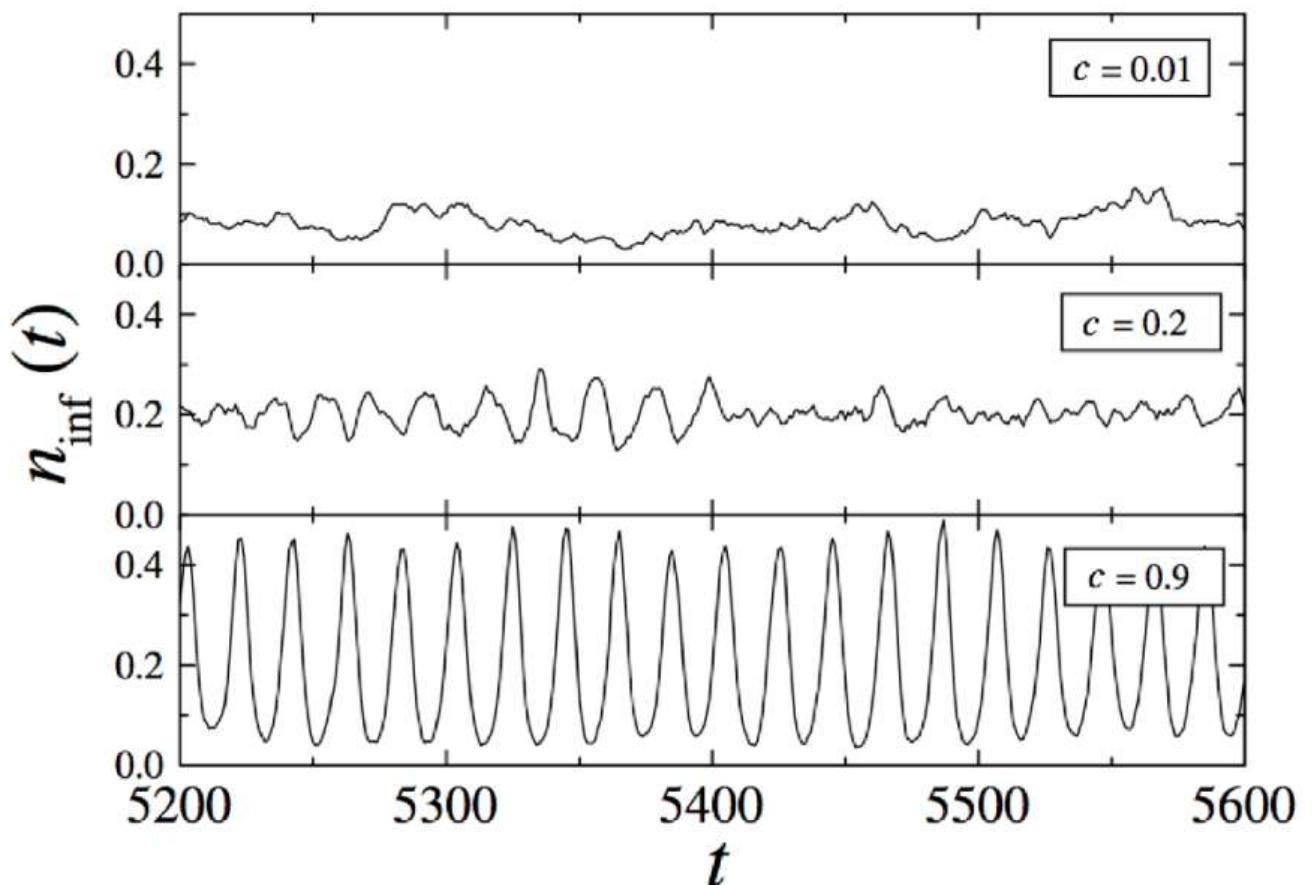
This means that we can be infected, after a while we recover and for a period of time we are immune to contracting the virus again. But in a couple of months or something like that, we can go back to being infected again, and this is the model that describes the Covid-19 the better.

108.1 Small World Contact Networks

The connection between Small World and the Synchronization/Oscillation is that temporary immunity can produce oscillations in very localized parts of the network, with patches of immunity following large numbers of infections in a concentrated area. This means that after a lot of people have been infected in a certain city, we can expect a large immunity since a lot of people have contracted the virus, and we have an oscillation since from a high number of infections, to a low number in the following months. But if the immunity continues to have this μ_r deadline, we can have the same process in a couple of years.

If we have many long range connections (such as continental flights) we introduce **coordination** with different places, since an infected person that is flying from turin to new york can start a synchronized contagious process in two very distant places in the world. We can build a **Small World network** following the Watts Strogatz model with nodes arranged in a ring, with many homophilous links, and some shortcuts, and we can add a probability c that controls the fraction of long-range links in the network.

While c is very low we don't have a significant synchronization, instead when c rises we observe the emergence of a synchronization process, and when c is quite high we can see that, in a period of time, we have a synchronization.



With the SIRS Model and the Small World contact network, where we have a small fraction of long range links in the network, while the c probability grows, the emergence of synchronization is much more likely.

109 ANALYSIS OF THE BRANCHING PROCESS

109.1 The Claim we want to Prove

We make some assumptions that aren't really that much real in the real world but it helps us to understand and to prove the claim

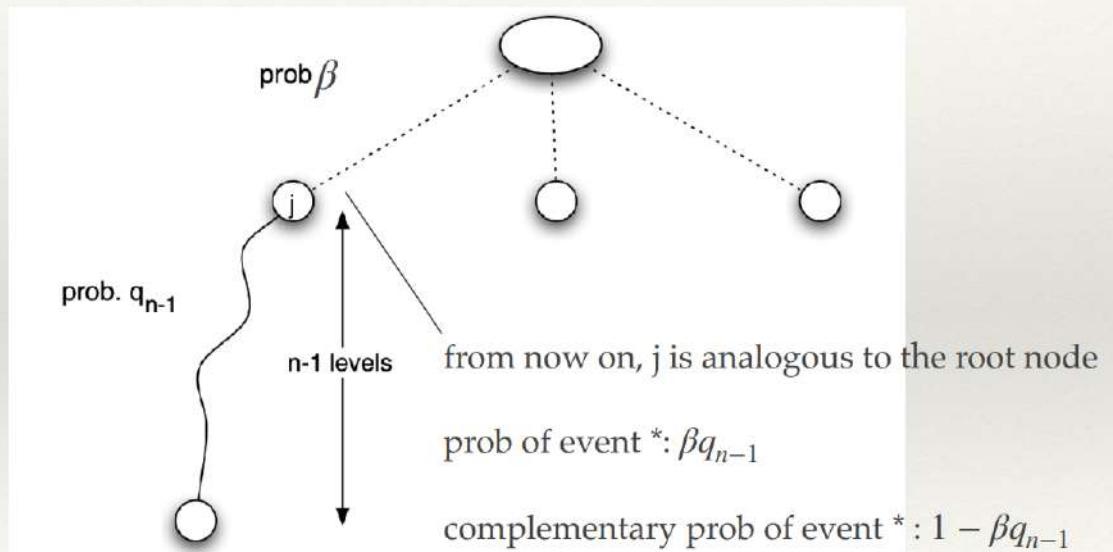
- ❖ The ingredients of our simple branching process:
 - ❖ contact network: an infinite tree
 - ❖ number of contacts: exactly k
 - ❖ contagion prob.: β
 - ❖ $R_0 = \beta k$
 - ❖ Let q_n the prob. that the epidemic **survives** for at least n waves, and $q^* = \lim_{n \rightarrow \infty} q_n$
- ❖ We want to prove the following
claim: (a) if $R_0 < 1$ then $q^* = 0$. (b) if $R_0 > 1$ then $q^* > 0$

We actually want to prove, in other words, is that if $R_0 < 1$ we can just sit back and let the epidemic do its course, since we know that it won't spread that much and become a pandemic. If $R_0 > 1$ then we know that the epidemic could and possibly will evolve into a pandemic.

Our goal is to prove the second part (b) of the claim and then the first (a). To prove this we want to find a formula for Q_n that can exactly determine its value.

We will try to write Q_n as a recursive function since it depends on k , beta and n and it is quite difficult to express a formula on these factors, hence we will write Q_n as Q_{n-1} . We can do this with the following reasoning:

Event *: The disease spreads through the root's node first contact j and then continues to persist down to n levels *in the part of the tree reachable through j*



Therefore, the disease fails to persist down to level n when **all the k copies** of event * fail to hold. This happens with prob.:

$$(1 - \beta q_{n-1})^k$$

We are almost done: this is also the probability $1 - q_n$, i.e., the probability that the disease fails to persist to n level

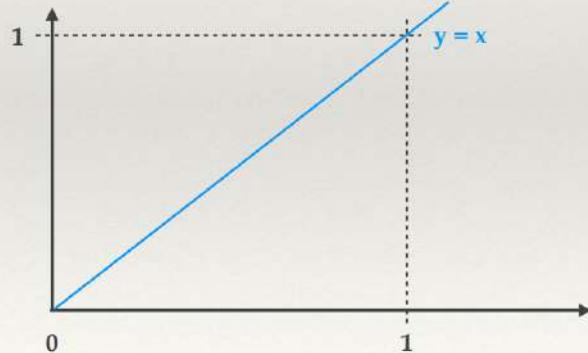
$$\begin{aligned} 1 - q_n &= (1 - \beta q_{n-1})^k \\ \Rightarrow q_n &= 1 - (1 - \beta q_{n-1})^k \end{aligned}$$

Now, assuming that the root is infected (patient 0): $q_0 = 1$

Finally, we can build up all the values q_1, q_2, q_3, \dots

109.2 Following the Values Q_n to a Limit

- We have a formula for $q_n = f(q_{n-1})$, where $f(x) = 1 - (1 - \beta x)^k$, we just want to follow the sequence: $1, f(1), f(f(1)), f(f(f(1))), \dots$ up to a limit
- Let's study the function $f(x) = 1 - (1 - \beta x)^k$ in the cartesian plane using some basic calculus tools



In the cartesian plane we have a function $y = x$ that divides the plane and we want to study this function

$$f(x) = 1 - (1 - \beta x)^k$$

$$f'(x) = \beta k (1 - \beta x)^{k-1}$$

$$f(0) = 0$$

$$0 \leq x \leq 1$$

$$f(1) = 1 - (1 - \beta)^k < 1$$

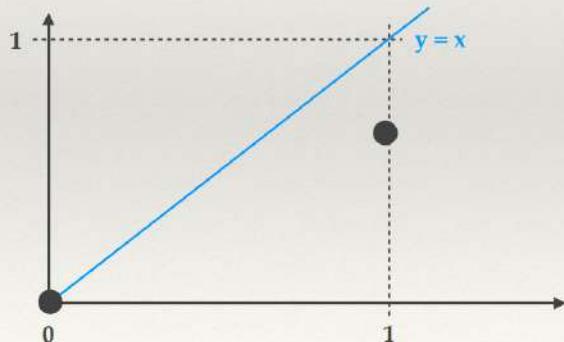
$$\Rightarrow f'(x) > 1 \wedge f'(0) > f'(1)$$

increasing and concave

$$f'(0) = \beta k = R_0$$

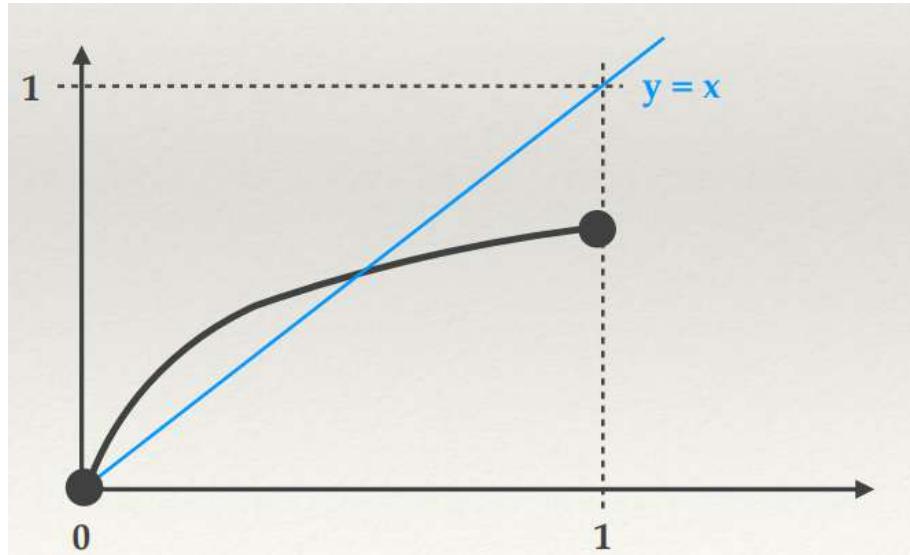
let's start with assuming
that: $R_0 > 1$

the slope of the tangent at
0 is > 1



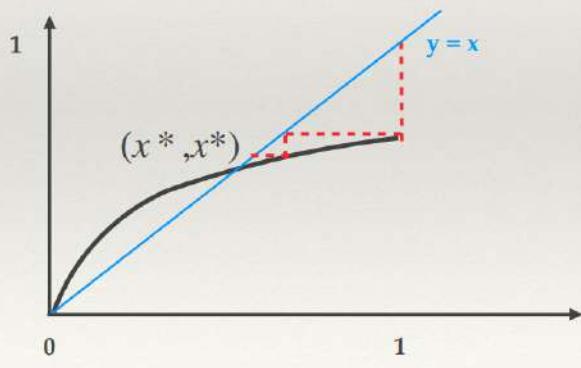
$f(0)$ is equal to 0 and we have our first point in the cartesian plane, $f(1)$ if we calculate it, we find out that it is a value lower than 1 and hence we put it a bit below 1. We now want to understand the behavior of the curve from the first to the second point and study it, and to do this we use the first derivative of the $f(x)$ formula and we need to study its behavior when x is between 0 and 1. As a result we find out that in the interval $0 \leq x \leq 1$ the derivative of x is greater than 1 and that $(f'(0)) > f'(1)$. This means that we have a curve that is increasing and concave.

When we are on this point we know that $f'(0)$ is equal to βk that is also equal to the basic reproductive number R_0 . We assume that $R_0 > 1$, so the second part of our claim, and that the slope of the tangent at 0 is greater than one.



As a result we have a curve that goes from 0 to 1, and when 0 is greater than 1, it intersects the straight function $y = x$.

- ❖ Recall that we want to follow the sequence: $1, f(1), f(f(1)), f(f(f(1))), \dots$ up to a limit
- ❖ We can track this sequence using the line $y = x$ (using the same technique of lecture 29)
- ❖ the process converges to the point (x^*, x^*) , where $y = x$ intersects $y = f(x)$



We start from a point in the cartesian plane that is $(1,1)$, and if we just move from the diagonal to the tip of the curve we find a point that is $(1, (f(1)))$, if we move again to the $y = x$ function we find $((f(1)), (f(1)))$ and so on and so forth. This is a geometrical process that allows us to always find the value of $f(1)$, and if we keep this process continuous we will stop at a point (x^*, x^*) where $y = x$ intersects $y = f(x)$.

- the sequence $1, f(1), f(f(1)), f(f(f(1))), \dots$ is precisely the sequence q_0, q_1, q_2, \dots , and it converges to $x^* > 0$, the unique point at which $f(x) = x$
- part (b) of the claim is proven: *if $R_0 > 1$ then $\lim_{n \rightarrow \infty} q_n > 0$*

This point means that the sequence of q_0, q_1, q_2 that we calculated recursively through f , converges to a point x^* that is greater than zero ($x > 0$), that is also the only point at which $f(x) = x$. We obtain this shape of the curve when $R_0 > 1$, and then the limit for n that tends to infinity of Q_n is greater than 0, and that means that we have a not neglectable probability that when $R_0 > 1$ the epidemic will survive at an indefinite time from now into the future and it could evolve into a global pandemic.

To prove the first part of the claim is quite easy, because we can see that if $R_0 < 1$, then the slope of the tangent at 0 is < 1 too and this would prove the first part of the claim that says that if $R_0 < 1$ then the limit for n that tends to infinity of Q_n is equal to zero. To see this we can just repeat the process that we did before for the second part of the proof in order to find the convergence, and we will obtain it when $x = 0$, and basically the limit is 0. **This says that when $R_0 < 1$ we can just rest assured that in the group that we are observing, the epidemic will vanish by itself.**

110 EPIDEMICS SPREADING IN HETEROGENEOUS NETWORKS

We have some information regarding model epidemic spreading that we need to recap in order to go further now which are

- ❖ We learnt how to model epidemic spreading
 - ❖ branching processes
 - ❖ compartmental models (i.e., SIR, SIS, SIRS)
- ❖ We found that the *basic reproductive number* $R_0 = \frac{\beta k}{\mu}$, a measure on how many people an infected individual can spread the disease, returns a useful **threshold**
 - ❖ if $R_0 < 1$, the epidemic will die out spontaneously
 - ❖ if $R_0 > 1$, the epidemic has a chance to survive forever
- ❖ Small changes on variables defining R_0 can have large effects on the spread of the disease

If we are observing a pandemic going on it is too late to stop the epidemic and even too late to calculate R_0 , so the only thing that we can do is to reduce the stress on the healthcare system by **flattening the curve**. The idea is to put on a cartesian plane the time since the first case and the number of cases, if we wait for the herd immunity we can expect a very high number of cases with lots of deaths and we want to keep this curve more flat as possible in order to not stress the health systems.

110.1 Real Networks

If we want to add more reality to our models, we need make some observations. We have made some assumptions that are not that real in the real life models and now we need to get rid of those:

- real networks are not trees as in the branching process
- we even made a **fully mixed hypothesis** assuming that each individual has the same chance to come into contact with an infected one and we know that this hypothesis is wrong, which basically means that we are assuming that people meet at random and hence we have homogeneously, but that is wrong and hence we need to introduce heterogeneity.

We need to give a little recap of **heterogeneity** to go further in this subject (NEXT PAGE)

❖ Let's define the **heterogeneity parameter** $\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$, where

$$\langle k \rangle = \frac{1}{N} \sum_i k_i \text{ that is the } \mathbf{mean}, \text{ and}$$

$$\langle k^2 \rangle = \frac{1}{N} \sum_i k_i^2, \text{ that is the } \mathbf{variance}$$

- ❖ This is a measure on how broad is the degree distribution
- ❖ In random networks: $\langle k^2 \rangle \approx \langle k \rangle^2 \Rightarrow \kappa \approx 1$
- ❖ Conversely, if the distribution is very heterogeneous: $\kappa \gg 1$

We also need to re-introduce the **Power Laws**

In many cases, the degree distribution is characterized by (an approx. of) a **power law function** $p_k \propto k^{-\gamma}$

It is possible to prove that when $2 < \gamma \leq 3$, then $\langle k^2 \rangle \rightarrow \infty$

That would imply that, with real data:

- (i) the variance will be different order of magnitude greater than the mean \Rightarrow the average has not statistical significance (**scale-free regime**)
- (ii) $\kappa \rightarrow \infty$, hence the distribution is highly heterogeneous

However, *power-law distributions* and *scale-free networks* are just particular cases of extreme heterogeneity; we want to focus on **heterogeneous networks** (that are very common in the real world)

In real life we have many hubs, and in real networks hubs are potentially super spreaders. If we are observing a cluster, it is very likely that in it we will find a super spreader hub, and this is a signal that the fully mixed hypothesis that every one meets the same amounts of people and hence contacts are not homogeneous. In order to understand what does it change if this hypothesis is removed and we will do it through a **continuum approach**.

111 A CONTINUUM APPROACH TO MODEL EPIDEMICS

We need revisit our modeling and propose a continuum approach that will transform the non deterministic model based on probability to some differential equations formulas that can help us to reason on what could happen in the long term.

111.1 Revisiting Comportmental Modeling

- ❖ To capture heterogeneity, we need to get rid of the fixed k contacts
- ❖ As a first approximation, we use an average $\langle k \rangle$
- ❖ An individual can be, as before, in one of these *states* : **S**, **I**, and **R**
- ❖ It also turns out that the best way to capture network dynamics is to move to a **continuous description of time t**
- ❖ $s(t)$, $i(t)$, and $r(t)$: the fraction of individuals who are susceptible, infectious, and removed at time t
- ❖ Initial conditions ($t=0$) - with only one infectious individual:
$$i(0) = \frac{1}{N}, s(0) = 1 - i(0) \quad \text{and, of course, } r(0) = 0$$

$i(0)$ is the number of total infected people, and $s(0)$ is the number of susceptible individuals which at $t = 0$ is just the whole 'world' minus the patient zero. And of course the number of recovered people at $t = 0$ is just 0.

111.2 SI Model

We consider the SI Model for Comportamental Modeling. In this model we don't consider recovery

- ❖ an infectious node transmit the diseases with prob. β : $i + s \xrightarrow{\beta} i + i$
- ❖ Let's assume there is no recovery and no one will be ever removed: when individuals enter the **I** compartment, they stay there forever
- ❖ Not hard to guess that everyone, when $t \rightarrow \infty$, will enter the infectious state.
- ❖ The question is: **how soon?**
- ❖ The plan: Let's find a formula for $i(t)$.
- ❖ *Full mixed hypothesis*: an infected individual comes into contact with other $\langle k \rangle s(t)$ susceptible individuals in a unit time

We want to understand the number of infectious individuals at time t . The idea is that the difference between the number of infected individuals from time t to time $t+1$ is given by the probability of meeting a susceptible individual and of being infected and multiplied by the transmission probability, and of course the number of susceptible individuals is going to decrease while the number of infected individuals is going to increase.

- ◊ This means that the new infections $di(t)$ during a time dt is:

$$di(t) = \beta\langle k \rangle s(t)i(t)dt$$

- ◊ observe that $s(t) = 1 - i(t)$. Moreover, for the sake of simplicity, let's get rid the (t) variable

$$\frac{di}{dt} = \beta\langle k \rangle i(1 - i)$$

- ◊ We can solve the above equation to obtain:

$$\frac{di}{i} + \frac{di}{1 - i} = \beta\langle k \rangle dt$$

- ◊ Integrating both sides:

$$\ln i - \ln(1 - i) + C = \beta\langle k \rangle t$$

- ◊ If we set a generic initial condition $i_0 = i(0)$, from the above equation:

$$\ln i - \ln(1 - i) + C = \beta\langle k \rangle t$$

- ◊ we get C :

$$C = \ln(1 - i_0) - \ln i_0$$

- ◊ Finally, after some algebraic substitutions:

$$\ln i - \ln(1 - i) + \ln(1 - i_0) - \ln i_0 = \beta\langle k \rangle t$$

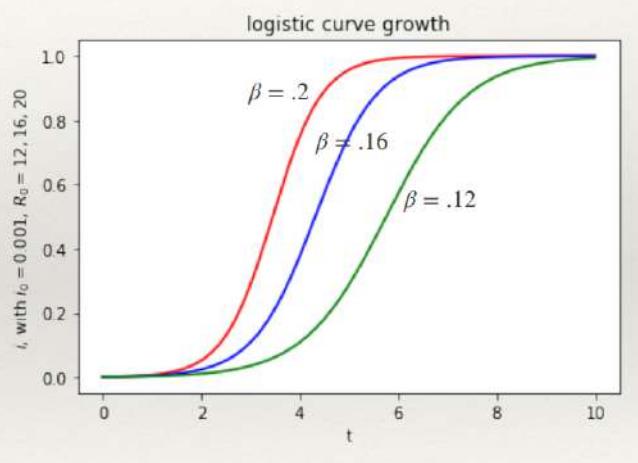
$$\Rightarrow \frac{i}{1 - i} \frac{1 - i_0}{i_0} = e^{\beta\langle k \rangle t} \Rightarrow \frac{1 - i_0}{i_0} i = e^{\beta\langle k \rangle t} - e^{\beta\langle k \rangle t} i$$

$$\Rightarrow \frac{1 - i_0 + i_0 e^{\beta\langle k \rangle t}}{i_0} i = e^{\beta\langle k \rangle t} \Rightarrow i = \frac{i_0 e^{\beta\langle k \rangle t}}{1 - i_0 + i_0 e^{\beta\langle k \rangle t}}$$

The point is that we obtained a formula that gives us deterministically an estimated number of the infected individuals at time t .

111.3 Logistic Growth Curve

- ❖ The equation $i = \frac{i_0 e^{\beta\langle k \rangle t}}{1 - i_0 + i_0 e^{\beta\langle k \rangle t}}$ is called *logistic growth curve*.
 - ❖ it grows exponentially at the beginning
 - ❖ After a while, there are very few susceptible around, and it is harder to infect some new individual
 - ❖ $\beta\langle k \rangle$ is the speed of the infection (no threshold!)



111.4 Observation of SI Epidemics

Very few diseases can be modeled with SI processes since the diseases that do not kill their victims, are eventually defeated by the immune system. SI models are still useful to understand that even the worst case scenario can be treated by delaying as much as possible the so called characteristic time tau (τ) which means you can move to the right the inflection point of the logistic curve. This time depends of Beta and heterogeneity parameter k, and is $\tau = 1 / \beta\langle k \rangle$. **We need to say that in this model there is no hope to avoid a global pandemic since sooner or later everyone will be infected, and R_0 is just a measure of how fast the pandemic is spreading.**

111.5 SIS Model

Can we find a similar formula for the SIS model as well? We need to say that in this model there is a possibility to recover with a certain amount of time, that is the recovery rate mu (μ), after which the individual can go back to the susceptible state.

- ❖ An infectious node transmit the diseases with prob. β : $i + s \xrightarrow{\beta} i + i$
- ❖ Infectious individuals recover with rate μ : $i \xrightarrow{\mu} s$
- ❖ New question: is there a threshold for $R_0 = \frac{\beta\langle k \rangle}{\mu}$?

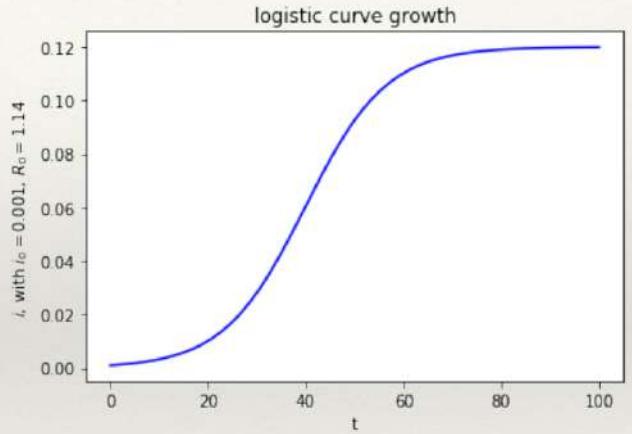
We can apply the same process that we have done for the SI model with just a couple of tweaks and we get the equation that describes this model. **We want to understand what happens to this functions when R0 changes and the time t goes to infinity.** In SI Models there is no hope, we will always observe an transformation that brings an epidemic to a pandemic, but we want to understand if R0 has the same role in the SIS model as in the SI model

- ❖ What happens to

$$i = \left(1 - \frac{\mu}{\beta\langle k \rangle}\right) \frac{Ce^{(\beta\langle k \rangle - \mu)t}}{1 + Ce^{(\beta\langle k \rangle - \mu)t}}$$

when $R_0 = \frac{\beta\langle k \rangle}{\mu}$ changes and $t \rightarrow \infty$?

- ❖ if $R_0 > 1 \Rightarrow \beta\langle k \rangle > \mu$,
then $\lim_{t \rightarrow \infty} i(t) = 1 - \frac{\mu}{\beta\langle k \rangle}$



When R_0 is greater than 1, the SIS model collapses into the SI model. This means that we can observe a slow start that exponentially grows very soon and then the curve will continue to grow but slowly compared to the first moments of the growth.

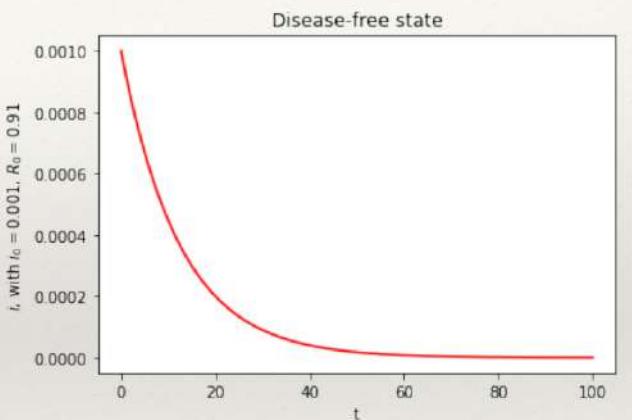
When R_0 is less than 1 and $\beta\langle k \rangle$ is less than μ , when t grows to infinity we see that the limit of the infected individuals goes to zero.

- ❖ What happens to

$$i = \left(1 - \frac{\mu}{\beta\langle k \rangle}\right) \frac{Ce^{(\beta\langle k \rangle - \mu)t}}{1 + Ce^{(\beta\langle k \rangle - \mu)t}}$$

when $R_0 = \frac{\beta\langle k \rangle}{\mu}$ changes and $t \rightarrow \infty$?

- ❖ if $R_0 < 1 \Rightarrow \beta\langle k \rangle < \mu$,
then $\lim_{t \rightarrow \infty} i(t) = 0$



This proves again that with the SIS models R_0 behaves as the typical rule of thumb way, and when it is lower than 1 or higher than 1 we have totally different scenarios.

112 SIR Model

We now want to understand if a certain disease could be better modeled with a SIR Model. Compared to the SIS model is that once we have recovered we are removed and not susceptible to infection any longer.

- ❖ An infectious node transmit the diseases with prob. β : $i + s \xrightarrow{\beta} i + i$
- ❖ Infectious individuals recover with rate μ : $i \xrightarrow{\mu} r$
- ❖ The question is: is there a threshold for $R_0 = \frac{\beta\langle k \rangle}{\mu}$ even here?

- ❖ The **equations** describing the dynamics of the **SIR model** are:

$$\frac{ds}{dt} = -\beta\langle k \rangle is = -\beta\langle k \rangle i(1 - r - i)$$

$$\frac{di}{dt} = \beta\langle k \rangle i(1 - r - i) - \mu i$$

$$\frac{dr}{dt} = \mu i$$

- ❖ and also: $s + r + i = 1$, $s(0) = s_0$, $r(0) = 0$, $i(0) = i_0 = 1 - s_0$

We have an equation that describes the dynamics of the SIR model that allows us to understand what happens to s , r , i . We know that the s , r , i individuals are equal to 1, we know that susceptible individuals at time 0 is equal to 0, and so for the recovered ones. While the infected people at time 0 are equal to the whole world 1 – the first susceptible patient at time 0.

We can study the values of R_0 by tweaking the parameters such as lowering the transmission rate by wearing masks or precautions, lowering k by imposing lockdowns, increasing or lowering the recovery rate μ by maybe administering certain medicines or such and we can see that by doing some of this things we can see that the curve is flattening and after a while the epidemic dies out. This is of course a very simple strategy that does not translate to the real world but it can help us to do some studies and develop some strategies.

112.1 Calculating the Asymptotic Value of Recovered Individuals r

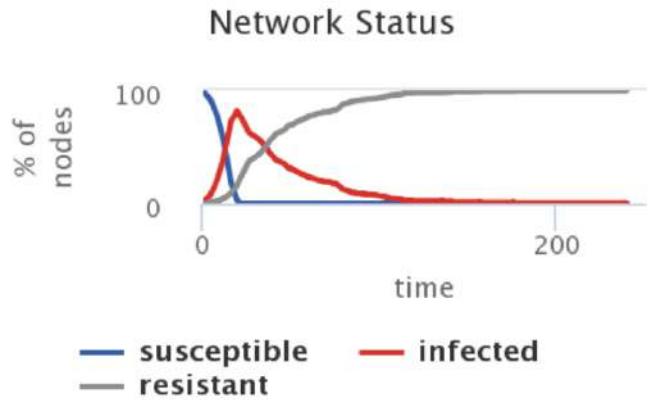
$$\frac{dr}{dt} = \mu(1 - r - s_0 e^{-R_0 r}) = 0$$

$$\Rightarrow r = 1 - s_0 e^{-R_0 r}$$

for large networks ($N \rightarrow \infty$): $s_0 \approx 1$

$$\Rightarrow r = 1 - e^{-R_0 r}$$

- ◊ if $R_0 > 1$,
then $\lim_{t \rightarrow \infty} r = 1$
- ◊ if $R_0 < 1$,
then $\lim_{t \rightarrow \infty} r = 0$



We find out that under this fully mixed approximations we see that R_0 has the same role for SIR and SIS.

113 HETEROGENEOUS NETWORKS EPIDEMICS

We now know that in a homogeneous scenario, R_0 is not only a measure of the speed of a certain epidemic or the expected peak of sick person, so it is not only a measure to understand what effects and tradeoffs we need to apply with the health care system in order to flatten the curve, but R_0 is a device that can tell us before hand if we need to worry or not.

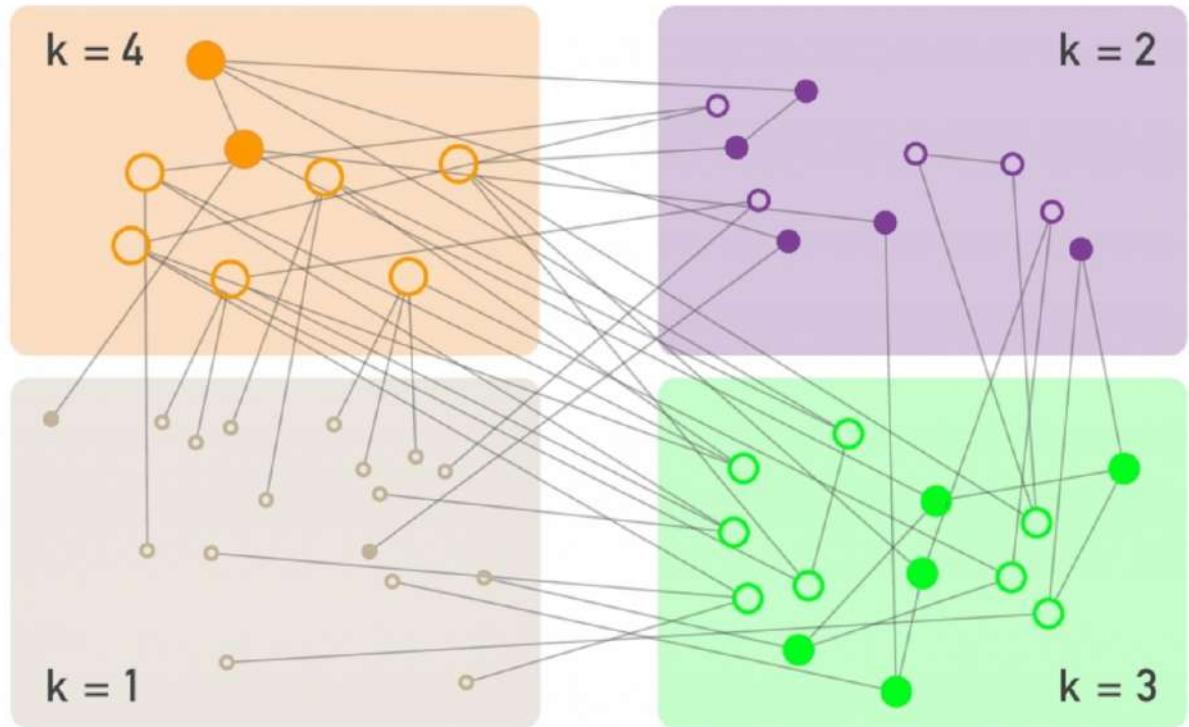
113.1 Removing the Full Mixed Hypothesis

We need to get rid of the full-mixed hypothesis because our networks have hubs, super-spreaders and are heterogeneous. To formalize this we need to introduce the **Degree-Block Approximation**.

113.2 Degree-Block Approximation

In heterogeneous networks, individuals with more links are more likely to be in contact with an infected individual, hence they are more likely to be infected. The degree-block approximation distinguishes nodes based on their degree and assumes that nodes with the same degree are statistically equivalent. Instead of calculating the fraction of infected individuals at time t , we calculate $i_k = I_k / N_k$ that is the fraction of nodes with degree k that are infected among all N nodes with degree k in the network.

$$i = \sum_k p_k i_k$$



We divide the population in classes of nodes with k degree, where k indicates the number of contacts of a certain node. The idea is that we can redefine all the differential equations for the SI model using the degree block approximation.

11.3.3 SI Model on a Network

- ❖ We write the SI model for each degree k separately

$$\frac{di_k}{dt} = \beta(1 - i_k)k\Theta_k$$
- ❖ Same structure as before, but
 - ❖ the average $\langle k \rangle$ is replaced by each node's actual degree k
 - ❖ we have the **density function** Θ_k : the fraction of infected neighbors of a susceptible node with degree k
 - ❖ now we have a system of k_{\max} coupled equations

The idea is that we can explore the early time behavior of the susceptible group of people with degree k .

- ❖ If we lack a cure, the only way to alter the course of an epidemic is to try to slow the spread of the epidemics as much as possible
- ❖ When i_k is small, the higher order term of $\beta i_k \Theta_k$ can be neglected:

$$\frac{di_k}{dt} \approx \beta k \Theta_k$$
- ❖ Assuming no degree correlation, we can prove that Θ_k is independent of k :

$$\Theta_k \approx i_0 \frac{\langle k \rangle - 1}{\langle k \rangle} e^{1/\tau}, \text{ where } \tau = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}$$
- ❖ τ is the *characteristic time* for the spread of the pathogen

The idea of this process is that we can plot the number of susceptible individuals in function of time and can still observe exponential growth at the early stage of the epidemic, so while k grows the characteristic time is close to 0. That means that if we observe that the group of nodes with higher degree has a higher fraction of infected nodes, as a consequence at any time all hubs are infected, but small-degree nodes tend to be disease free. Hence the disease is maintained in the hubs, which in turn broadcast the disease to the rest of the network.

113.4 Observation on the Characteristic Time τ

- ❖ The characteristic time $\tau = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}$ depends on the heterogeneity of the network
- ❖ In **random networks**, i.e., $\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle^2} \approx 1 \Rightarrow \langle k^2 \rangle \approx \langle k \rangle^2 \Rightarrow \tau \approx \frac{1}{\beta \langle k \rangle}$
 (recovering the results under homogeneous assumption)
- ❖ In **heterogeneous networks**, i.e., $\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle^2} \gg 1 \Rightarrow \langle k^2 \rangle \gg \langle k \rangle^2 \Rightarrow \tau \ll \frac{1}{\beta \langle k \rangle}$
- ❖ In **scale-free networks** (special case) we have that $\langle k^2 \rangle \xrightarrow{N \rightarrow \infty} \infty \Rightarrow \tau \rightarrow 0$

We already knew that with SI dynamics the diseases would reach all the individual and we use characteristic time to understand the speed of the diffusion. In homogeneous networks $\tau = 1 / \beta \langle k \rangle$ and some countermeasures can delay the propagation, while in heterogeneous networks the spread is almost instantaneous because characteristic time goes to 0 very early.

113.5 SIS Model in a Network

- ❖ To understand the full impact of the network topology, let's explore the SIS model further
- ❖ The continuum equation describing the dynamics of the extension of the SIS model on a network is the following:

$$\frac{di_k}{dt} = \beta(1 - i_k)k\Theta_k(t) - \mu i_k$$

- ❖ The presence of the recovery term $-\mu i$ changes the characteristic time to:

$$\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \mu \langle k \rangle}$$

After defining the continuum equation that describes the SIS model we can define an **Epidemic Threshold**.

113.6 Epidemic Threshold

- ❖ Let's define $\lambda = \frac{\beta}{\mu}$ as the **spreading rate**: it depends only on biological characteristics of the pathogen
 - ❖ Note that $R_0 = \lambda \langle k \rangle$
 - ❖ The higher is λ , the more likely that the diseases will spread
- ❖ Let's look for the **epidemic threshold** λ_c : the pathogen can spread only if $\lambda > \lambda_c$

- Recall that $\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \mu \langle k \rangle}$, we want to find when the pathogen has a chance to persist:

$$\tau > 0 \Rightarrow \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \mu \langle k \rangle} > 0 \Rightarrow \beta \langle k^2 \rangle > \mu \langle k \rangle \Rightarrow \lambda = \frac{\beta}{\mu} > \frac{\langle k \rangle}{\langle k^2 \rangle};$$

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

As a result we have that:

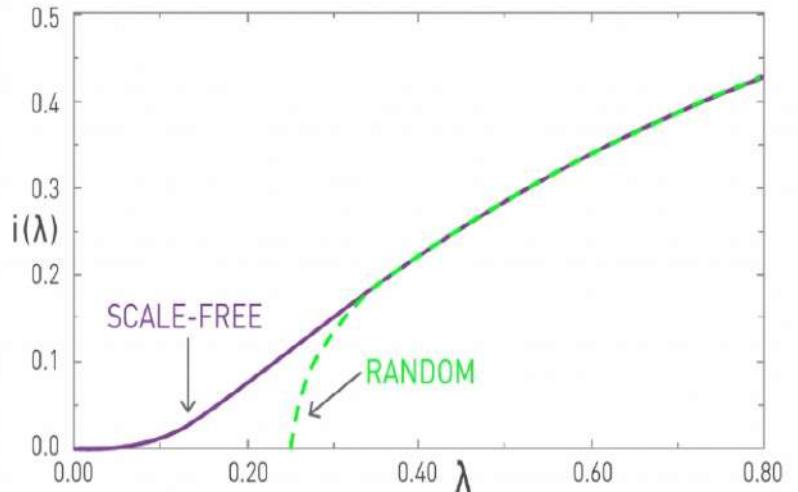
- In **random networks**, i.e., $\langle k^2 \rangle \approx \langle k \rangle^2 \Rightarrow \lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle} \approx \frac{1}{\langle k \rangle}$ that confirms that, under homogeneous hypothesis, the pathogen has a chance to persist when $\lambda > \frac{1}{\langle k \rangle}$, or analogously, when $R_0 > 1$
- In **heterogeneous networks**, i.e., $\langle k^2 \rangle \gg \langle k \rangle^2 \Rightarrow \lambda_c$ is expected to vanish for increasingly heterogeneous networks
- In **scale-free networks** (special case), we have that $\langle k^2 \rangle \xrightarrow{N \rightarrow \infty} \infty \Rightarrow \lambda_c \rightarrow 0$

ENDING ON THE NEXT PAGE

The random network has a finite epidemic threshold λ_c , implying that a pathogen with a small spreading rate ($\lambda < \lambda_c$) must die out, i.e. $i(\lambda) = 0$.

If, however, the spreading rate of the pathogen exceeds λ_c , the pathogen becomes **endemic** and a finite fraction of the population is infected at any time.

For a **scale-free network** we have $\lambda_c = 0$, hence even viruses with a very small spreading rate λ can persist in the population.



R_0 is still useful, but we cannot rely on it being greater or lower than 1 to understand how the epidemic will behave.

113.7 No Epidemic Threshold and Hubs

We can have similar results also with SIR models, we have **vanishing threshold in heterogeneous networks**, and a direct consequence of hubs is that even a weakly infectious virus can spread widely in a population through **hub super-spreaders**. If the virus fails to infect a hub, it can die out very soon.

One way to reduce the speed of infections is to use immunizations strategies to lower the spreading power of hubs, using vaccines, medicines etc..