Algorithm Foundations of Data Science

Lecture 3: Graph and Patterns

MING GAO

DaSE@ECNU
(for course related communications)
mgao@dase.ecnu.edu.cn

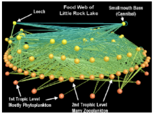
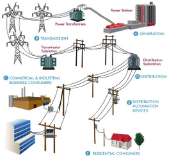
Mar. 28, 2018

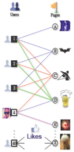
Outline

Outline

Graphs - why should we care?





Networks in real world

- "YahooWeb graph": 1B vertices(Web sites), 6B edges (http links)
- Facebook, Twitter, etc: more than 1B users
- Food Web: all biologies, food chain
- Power-grid: vertices (plants or consumers), edges (power lines)
- Airline route: vertices (airports), edges (flights)
- Adoption: users purchase products, adopt services, etc.

Motivation questions

Questions

- What do real graphs look like?

Motivation questions

Questions

- What do real graphs look like?
  - What properties of vertices, edges are important to model?
  - What local and global properties are important to measure?

Motivation questions

Questions

- What do real graphs look like?
    - What properties of vertices, edges are important to model?
    - What local and global properties are important to measure?
- Are graphs helpful to understand the real world?
    - Social influence
    - Recommendation
    - Information propagation
    - Human behaviors

Motivation questions

Questions

- What do real graphs look like?
  - What properties of vertices, edges are important to model?
  - What local and global properties are important to measure?
- Are graphs helpful to understand the real world?
  - Social influence
  - Recommendation
  - Information propagation
  - Human behaviors
- Is a sub-graph “normal” (Water army, fraud detection, spam filtering, etc)?

Motivation questions

Questions

- What do real graphs look like?
  - What properties of vertices, edges are important to model?
  - What local and global properties are important to measure?
- Are graphs helpful to understand the real world?
  - Social influence
  - Recommendation
  - Information propagation
  - Human behaviors
- Is a sub-graph "normal" (Water army, fraud detection, spam filtering, etc)?
- How to generate realistic graphs?

Motivation questions

Questions

- What do real graphs look like?
  - What properties of vertices, edges are important to model?
  - What local and global properties are important to measure?
- Are graphs helpful to understand the real world?
  - Social influence
  - Recommendation
  - Information propagation
  - Human behaviors
- Is a sub-graph "normal" (Water army, fraud detection, spam filtering, etc)?
- How to generate realistic graphs?
- How to get a "good" sample of a network?

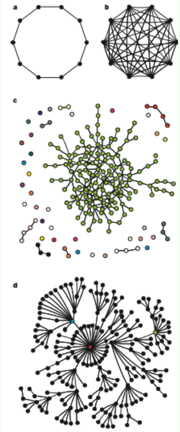
Motivation questions

Questions

- What do real graphs look like?
  - What properties of vertices, edges are important to model?
  - What local and global properties are important to measure?
- Are graphs helpful to understand the real world?
  - Social influence
  - Recommendation
  - Information propagation
  - Human behaviors
- Is a sub-graph "normal" (Water army, fraud detection, spam filtering, etc)?
- How to generate realistic graphs?
- How to get a "good" sample of a network?
- How to design an efficient algorithm to handle large-scale graphs?

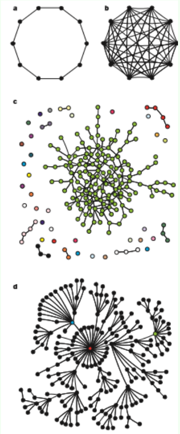
Models for complex networks

Steven H. S. proposes the model for complex networks in Nature 2001.



Models for complex networks

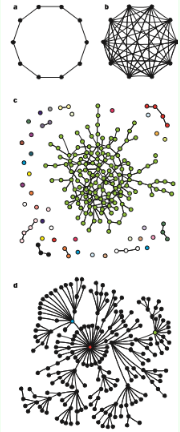
Steven H. S. proposes the model for complex networks in Nature 2001.



- Regular network: each node has exactly the same number of edges.

Models for complex networks

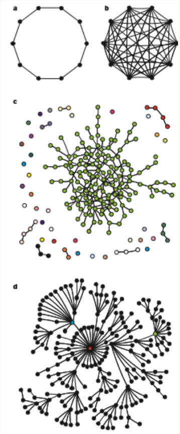
Steven H. S. proposes the model for complex networks in Nature 2001.



- Regular network: each node has exactly the same number of edges.
- Random network: it is obtained by starting with a set of n isolated vertices and adding successive edges between them at random.

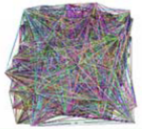
Models for complex networks

Steven H. S. proposes the model for complex networks in Nature 2001.

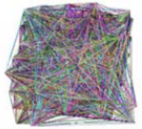


- Regular network: each node has exactly the same number of edges.
- Random network: it is obtained by starting with a set of n isolated vertices and adding successive edges between them at random.
- Scale-free network: it grows via attaching new nodes to previously existing nodes randomly, while the probability is proportional to the degree of the target node, i.e., richly connected nodes tend to get richer, leading to the formation of hubs and a skewed degree distribution with a heavy tail.(Matthew Effect or Pareto's Law)

Are real graphs random?



Are real graphs random?



Looks random - right?

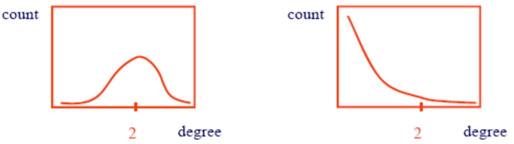
How does the Internet look like? Any rules?

Are real graphs random?

Looks random - right?

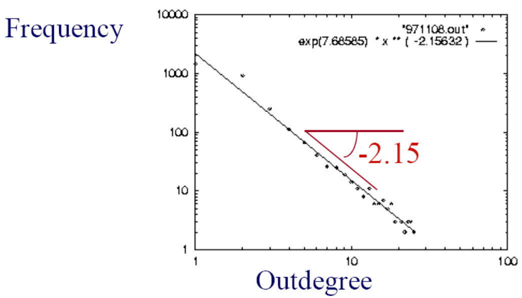
How does the Internet look like? Any rules?

- Diameter: would you like to guess?
- In- and out- degree distributions: if average degree is 2, what is the most probable degree?
- Other (surprising) patterns?

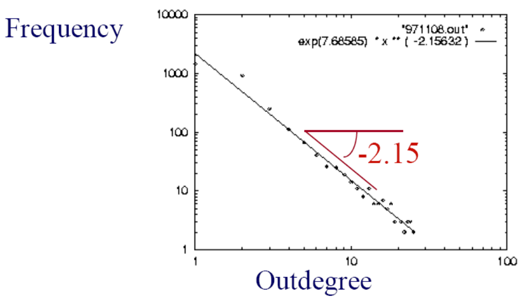


Outline

Power-law I


Frequency
10000
1000
100
10
1
'971108.out'
exp(7.68585) * x ** (-2.15632)
-2.15
1
10
100
Outdegree

Power-law I



Internet topology

- Out-degree distribution is plotted in log-log scale.
- It forms a line with a slope ~ -2.15
- $freq. = deg.^{-2.15}$

What is the power-law?

Due to Matthew effect, Pareto's law, "rich-get-richer", or the 80/20 principle, there are many settings with power law (Zipf's law).

What is the power-law?

Due to Matthew effect, Pareto's law, "rich-get-richer", or the 80/20 principle, there are many settings with power law (Zipf's law).

- 80% of Italy's land owned by 20% of the population.

What is the power-law?

Due to Matthew effect, Pareto's law, "rich-get-richer", or the 80/20 principle, there are many settings with power law (Zipf's law).

- 80% of Italy's land owned by 20% of the population.
- Richest 20% obtain 82.70% income.

What is the power-law?

Due to Matthew effect, Pareto's law, "rich-get-richer", or the 80/20 principle, there are many settings with power law (Zipf's law).

- 80% of Italy's land owned by 20% of the population.
- Richest 20% obtain 82.70% income.
- Bible: rank VS. frequency (log-log)

What is the power-law?

Due to Matthew effect, Pareto's law, "rich-get-richer", or the 80/20 principle, there are many settings with power law (Zipf's law).

- 80% of Italy's land owned by 20% of the population.
- Richest 20% obtain 82.70% income.
- Bible: rank VS. frequency (log-log)
- Web: hit count VS. volume

What is the power-law?

Due to Matthew effect, Pareto's law, "rich-get-richer", or the 80/20 principle, there are many settings with power law (Zipf's law).

- 80% of Italy's land owned by 20% of the population.
- Richest 20% obtain 82.70% income.
- Bible: rank VS. frequency (log-log)
- Web: hit count VS. volume
- File: count VS. size

What is the power-law?

Due to Matthew effect, Pareto's law, "rich-get-richer", or the 80/20 principle, there are many settings with power law (Zipf's law).

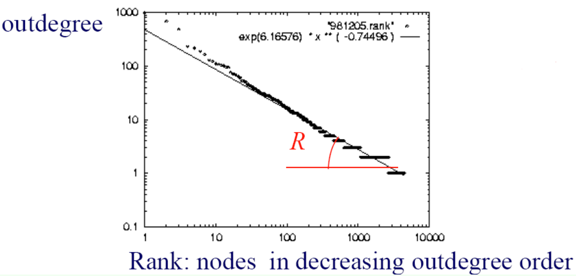
- 80% of Italy's land owned by 20% of the population.
- Richest 20% obtain 82.70% income.
- Bible: rank VS. frequency (log-log)
- Web: hit count VS. volume
- File: count VS. size
- Publication: citation VS. count

What is the power-law?

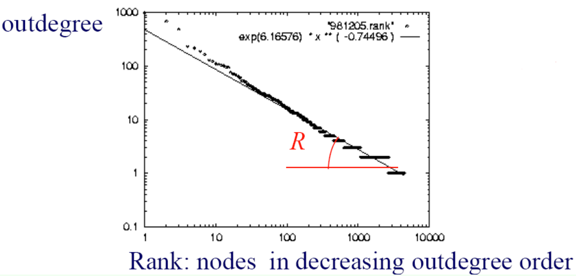
Due to Matthew effect, Pareto's law, "rich-get-richer", or the 80/20 principle, there are many settings with power law (Zipf's law).

- 80% of Italy's land owned by 20% of the population.
- Richest 20% obtain 82.70% income.
- Bible: rank VS. frequency (log-log)
- Web: hit count VS. volume
- File: count VS. size
- Publication: citation VS. count
- Business
  - 80% of a company's profits come from 20% of customers.
  - 80% of a company's complaints come from 20% of customers.
  - 80% of a company's profits come from 20% of the time staff spent
  - 80% of a company's sales are made by 20% of sales staff

Power-law II


outdegree
"981205.rank"
exp(6.16576) * x ** (-0.74496)
R
1000
100
10
1
0.1
1
10
100
1000
10000
Rank: nodes in decreasing outdegree order

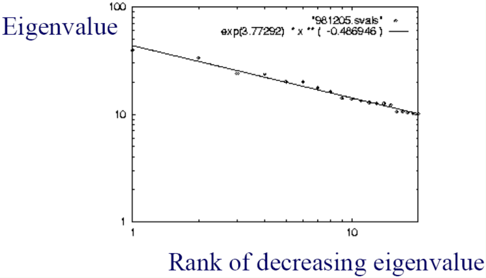
Power-law II



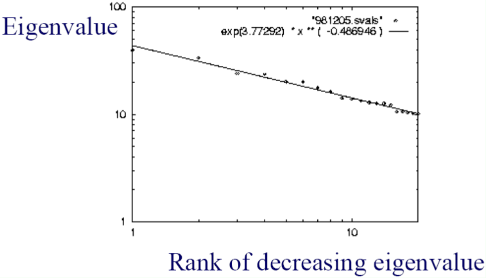
Rank of out-degrees

- Vertices are ranked in decreasing out-degree order, and plotted in log-log scale.
- It forms a line with a slope ~ -0.74
- $deg. = rank^{-0.74}$

Power-law III



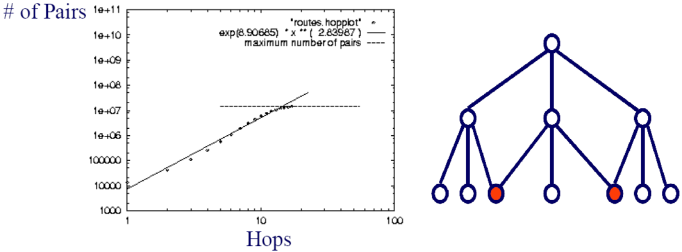
Power-law III



Rank of eigenvalues

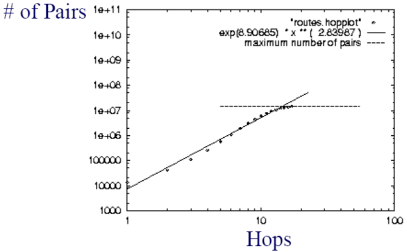
- Eigenvalues of adjacency matrix (top 20) are ranked in decreasing order, and plotted in log-log scale.
- It forms a line with a slope ~ -0.48
- $eigen. = rank^{-0.48}$

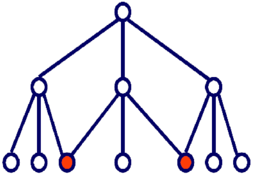
Power-law IV



# of Pairs
Hops
"routes.hopplot"
exp(8.90685) * x ** (2.83987)
maximum number of pairs
1e+11
1e+10
1e+09
1e+08
1e+07
1e+06
100000
10000
1000
1
10
100

Power-law IV

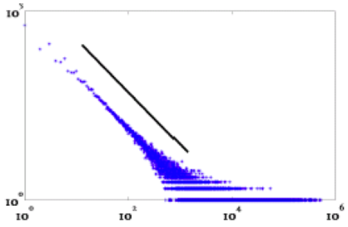




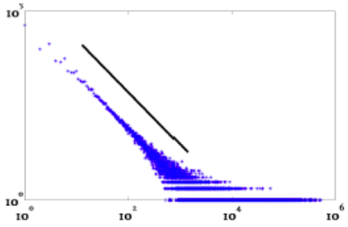
Hop plot

- How many neighbors within $1, 2, \cdots, h$ hops? $(\sum_{i=1}^{h} avg.^{i})$
- Pairs of vertices are plotted in log-log scale. It forms a line with a slope ~ 2.83
- $pairs. = hop^{2.83}$

Power-law V



Power-law V



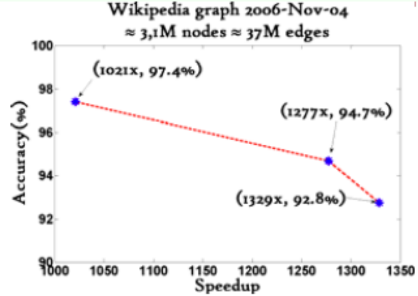
Counting of triangles

- X-axis: # of triangles a vertex participates in
- Y-axis: count of such vertices
- In log-log scale, the plot is almost linear.

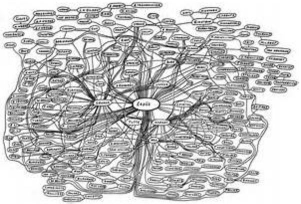
Triangle law

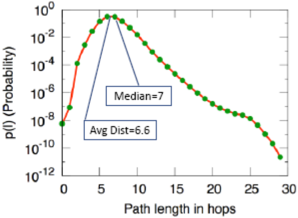
How to count # triangles?

- Naive algorithm: 3-way join ($O(n^3)$).
- # triangles $= \frac{1}{6}\sum_{i=1}^{n}\lambda_i^3$. Why?
- Because of skewness, we only need the top few eigenvalues via using Lanczos algorithm.

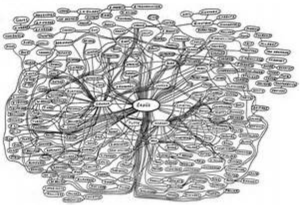


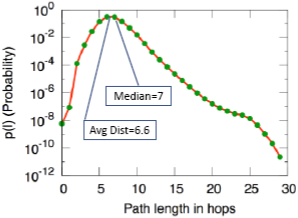
Erdös number





Erdös number



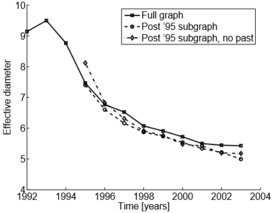


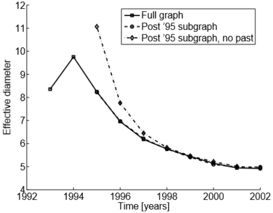
Small world - six degrees of separation

The world looks “small” when you think of how short a path of friends it takes to get from you to almost anyone else. Stanley Milgram and his colleagues in the 1960s did an experiment.

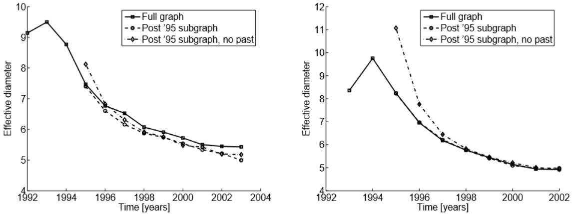
- 296 randomly chosen “starters” asked to forward a letter to a “target” person, a stockbroker in Boston’s suburb.
- The six degrees of separation was also found by Jure Leskovec on Miscrosoft Instant Message.

Shrinking diameter





Shrinking diameter



Citation or patents networks

For citation network, they collected citations among Physics papers.

- 11 years data
  - 29,555 papers
  - 352,807 citations
- For each month, create a graph of all citations up to the month.
- The diameters are plotted in the figures.

Temporal evolution of graphs

Question

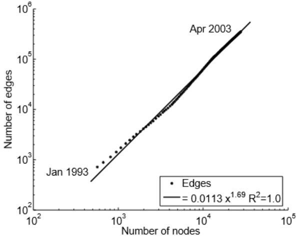
Let $N(t)$ and $E(t)$ be # nodes and # edges at time t, respectively. Suppose that $N(t+1) = 2N(t)$, what is your guess for $E(t+1)$?

Temporal evolution of graphs

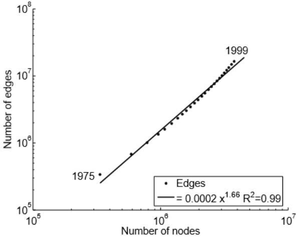
Question

Let $N(t)$ and $E(t)$ be # nodes and # edges at time t, respectively. Suppose that $N(t+1) = 2N(t)$, what is your guess for $E(t+1)$?

- It is over-doubled, but obeying: $E(t) \sim N(t)^{\alpha}$ for all t, where $1 < \alpha < 2$.
- For tree (clique), $\alpha = 1$ ($\alpha = 2$).



(a) arXiv



(b) Patents

Dunbar's number

Why primates have unusually big brains?

Social group size (and a lot of social behaviour as wel) correlates with relative neocortex volume.

Dunbar's number

Why primates have unusually big brains?

Social group size (and a lot of social behaviour as wel) correlates with relative neocortex volume.

- Our relationships form a hierarchically inclusive series of circles of increasing size but decreasing intensity.

Dunbar's number

Why primates have unusually big brains?

Social group size (and a lot of social behaviour as wel) correlates with relative neocortex volume.

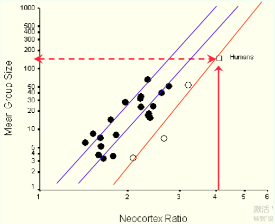
- Our relationships form a hierarchically inclusive series of circles of increasing size but decreasing intensity.
- 150 is the limitation on reciprocated relationships.

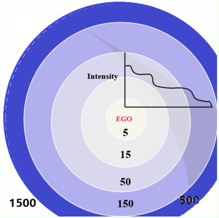
Dunbar's number

Why primates have unusually big brains?

Social group size (and a lot of social behaviour as wel) correlates with relative neocortex volume.

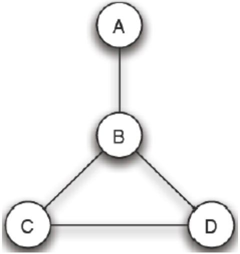
- Our relationships form a hierarchically inclusive series of circles of increasing size but decreasing intensity.
- 150 is the limitation on reciprocated relationships.
- 1500 is the limitation on memory for faces?





Outline

Graph types

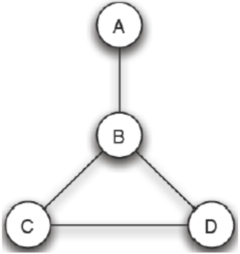


Undirected graph

A undirected graph on 4 vertices

- Degree: # edges connected to the vertex
- Degree 0 vertex: isolated vertex

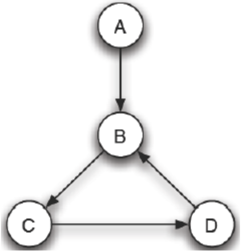
Graph types



Undirected graph

A undirected graph on 4 vertices

- Degree: # edges connected to the vertex
- Degree 0 vertex: isolated vertex

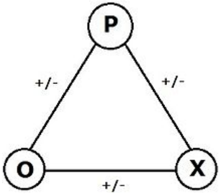


Directed graph

A directed graph on 4 vertices

- In-degree: # incoming edges to the vertex
- Out-degree: # outgoing edges to the vertex
- Degree: in-degree + outdegree

Graph types cont.

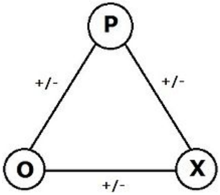


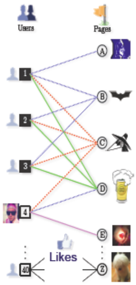
Signed graph

A signed graph on 3 vertices

- Positive-degree: # edges associated with positive labels
- Negative-degree: # edges associated with negative labels

Graph types cont.





Signed graph

A signed graph on 3 vertices

- Positive-degree: # edges associated with positive labels
- Negative-degree: # edges associated with negative labels

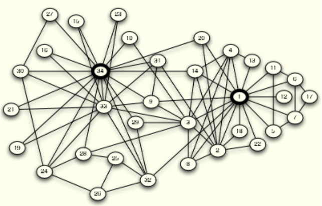
Bipartite graph

Users interact on social platforms

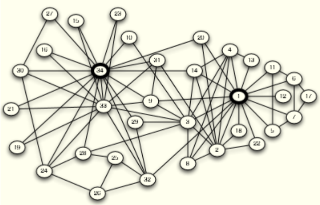
- Reply network
- Retweet network
- Adoption network

Outline

Paths


27
15
23
10
20
16
4
13
31
11
30
34
14
6
1
12
17
21
33
9
29
3
7
5
18
19
28
22
25
2
24
8
32
26

Paths

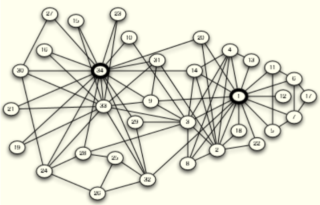


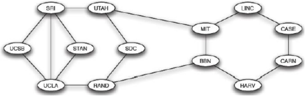
Path

Path is a sequence of nodes with the property that each consecutive pair in the sequence is connected by an edge

- Simple path does not repeat nodes.
- The length of path is the number of nodes in the path

Paths



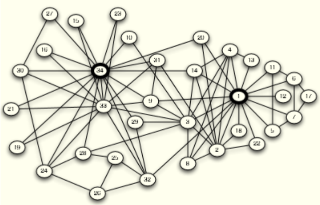


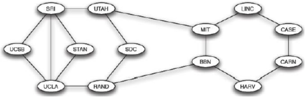
Path

Path is a sequence of nodes with the property that each consecutive pair in the sequence is connected by an edge

- Simple path does not repeat nodes.
- The length of path is the number of nodes in the path

Paths





Path

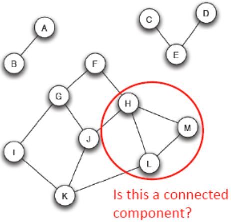
Path is a sequence of nodes with the property that each consecutive pair in the sequence is connected by an edge

- Simple path does not repeat nodes.
- The length of path is the number of nodes in the path

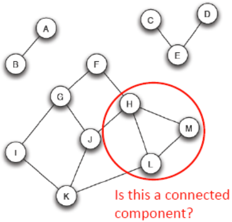
Cycle

Cycle is a path with at least three edges, in which the first and last nodes are the same. Every edge in the 1970 Arpanet belongs to a cycle, and this was by design. Why?

Connectivity



Connectivity

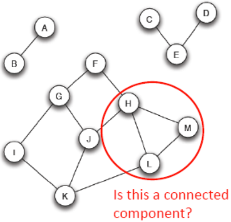


Connected component

A connected component is a subset of nodes s.t.:

- Every node in the subset has a path to every other; and

Connectivity

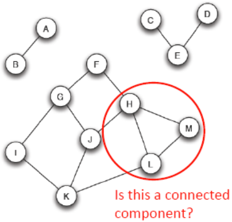


Connected component

A connected component is a subset of nodes s.t.:

- Every node in the subset has a path to every other; and
- The subset is not part of some larger set with the property that every node can reach every other.

Connectivity



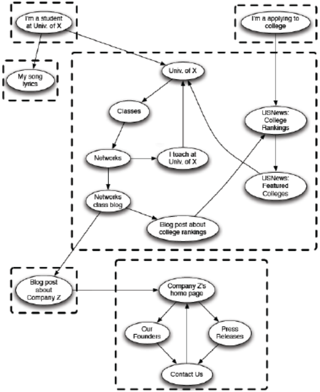
Connected component

A connected component is a subset of nodes s.t.:

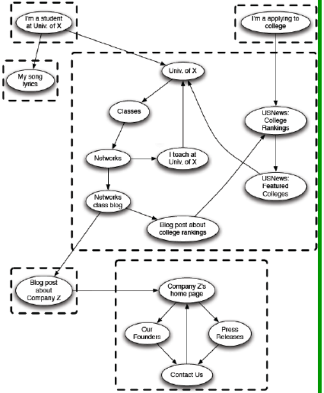
- Every node in the subset has a path to every other; and
- The subset is not part of some larger set with the property that every node can reach every other.

A graph is connected if for every pair of nodes, there is a path between them, i.e., the whole graph is a connected component.

Strongly connected component



Strongly connected component

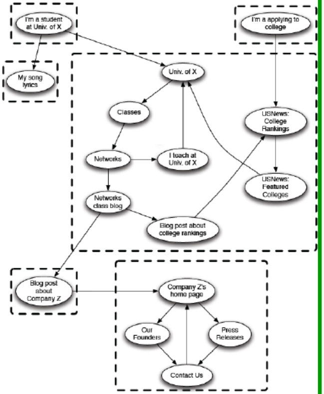


Strongly connected component

A *directed graph* is strongly connected if there is a path from every node to every other node.

- Edges of the path must follow the forward direction.

Strongly connected component

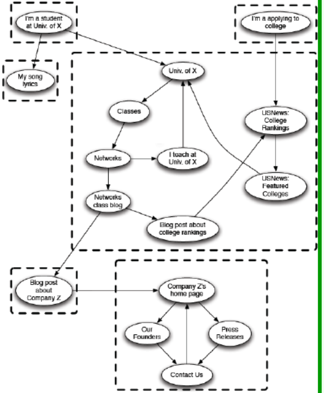


Strongly connected component

A *directed graph* is strongly connected if there is a path from every node to every other node.

- Edges of the path must follow the forward direction.
- A undirected graph can be treated as a bidirectional graph. Thus connected component in a directed graph is also a SCC.

Strongly connected component

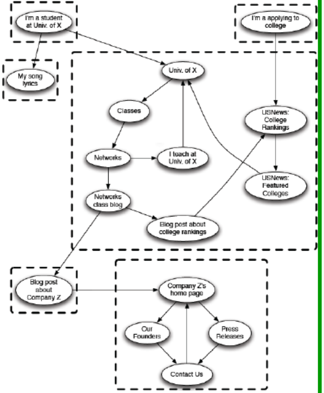


Strongly connected component

A *directed graph* is strongly connected if there is a path from every node to every other node.

- Edges of the path must follow the forward direction.
- A undirected graph can be treated as a bidirectional graph. Thus connected component in a directed graph is also a SCC.
- In a strongly connected component, there are followers and followees for each node.

Strongly connected component

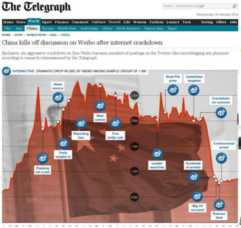


Strongly connected component

A *directed graph* is strongly connected if there is a path from every node to every other node.

- Edges of the path must follow the forward direction.
- A undirected graph can be treated as a bidirectional graph. Thus connected component in a directed graph is also a SCC.
- In a strongly connected component, there are followers and followees for each node.
- SCCs can be treated as super-nodes.

Giant component



The Telegraph

Wednesday 19 February 2014

Home News World Sport Finance Comment Culture Travel Life Women Fashion Luxury Tech Dating Offers Jobs

USA Asia China Europe Middle East Australasia Africa South America Central Asia Expat

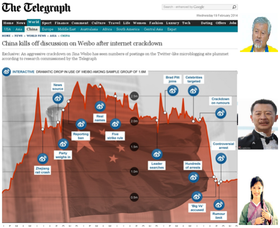
HOME » NEWS » WORLD NEWS » ASIA » CHINA

China kills off discussion on Weibo after internet crackdown

Exclusive: An aggressive crackdown on Sina Weibo has seen numbers of postings on the Twitter-like microblogging site plummet according to research commissioned by the Telegraph

INTERACTIVE DRAMATIC DROP IN USE OF WEIBO AMONG SAMPLE GROUP OF 1.6M

Giant component

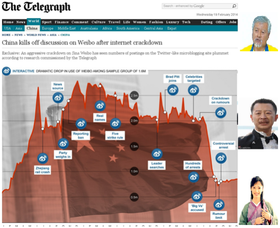


Giant connected component

A connected component that contains a significant fraction of all the nodes.

- When a network (e.g., friendship network) contains a giant component, it almost always contains only one.

Giant component

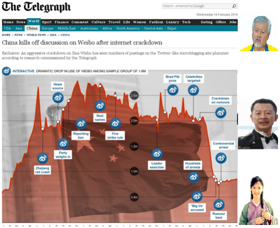


Giant connected component

A connected component that contains a significant fraction of all the nodes.

- When a network (e.g., friendship network) contains a giant component, it almost always contains only one.
- The other connected components are very small by comparison.

Giant component

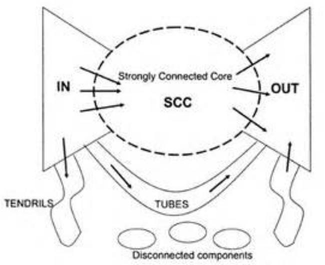


Giant connected component

A connected component that contains a significant fraction of all the nodes.

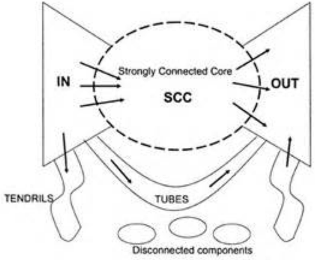
- When a network (e.g., friendship network) contains a giant component, it almost always contains only one.
- The other connected components are very small by comparison.
- The largest connected component would break apart into three distinct components if this node were removed [related to robustness of network].

Web giant component



200 M pages, 1.5 B hyperlinks

Web giant component

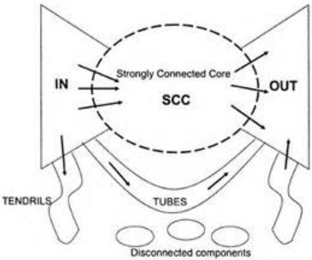


200 M pages, 1.5 B hyperlinks

Web graph

Web contains a giant strongly connected component (containing home pages of many of the major commercial, governmental, and non-profit organizations)

Web giant component



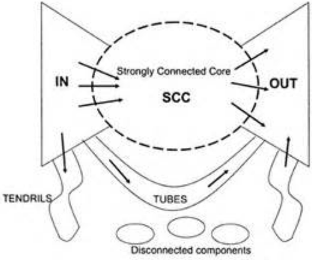
200 M pages, 1.5 B hyperlinks

Web graph

Web contains a giant strongly connected component (containing home pages of many of the major commercial, governmental, and non-profit organizations)

- IN: nodes that can reach the giant SCC but cannot be reached from it, i.e., nodes that are "upstream" of it.

Web giant component



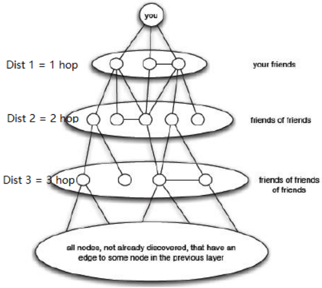
200 M pages, 1.5 B hyperlinks

Web graph

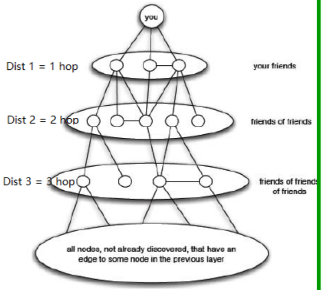
Web contains a giant strongly connected component (containing home pages of many of the major commercial, governmental, and non-profit organizations)

- IN: nodes that can reach the giant SCC but cannot be reached from it, i.e., nodes that are "upstream" of it.
- OUT: nodes that can be reached from the giant SCC but cannot reach it, i.e., nodes are "downstream" of it.

Distance and diameter



Distance and diameter

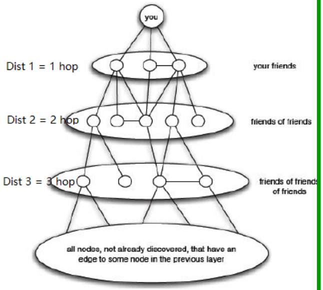


Distance or Geodesic distance

The distance between two vertices in a graph is the number of edges in a shortest path.

- Diameter is the length of the "longest shortest path" between any two vertices of a graph.

Distance and diameter

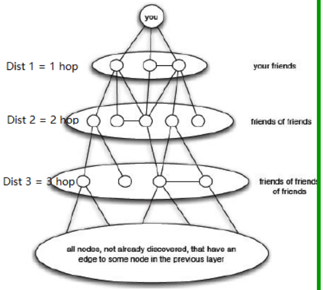


Distance or Geodesic distance

The distance between two vertices in a graph is the number of edges in a shortest path.

- Diameter is the length of the "longest shortest path" between any two vertices of a graph.
- Erdös number is bounded by diameter of a graph.

Distance and diameter



Distance or Geodesic distance

The distance between two vertices in a graph is the number of edges in a shortest path.

- Diameter is the length of the "longest shortest path" between any two vertices of a graph.
- Erdös number is bounded by diameter of a graph.
- Research community is a small world [Duncan Watts and Steven Strogatz 1998].

Mean Geodesic distance of undirected networks

Definition

$$L = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij},$$

where n denotes # of nodes, and d_{ij} is the shortest distance between nodes i and j.

- Mean Geodesic distance includes distance to itself.

Mean Geodesic distance of undirected networks

Definition

$$L = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij},$$

where n denotes # of nodes, and d_{ij} is the shortest distance between nodes i and j.

- Mean Geodesic distance includes distance to itself.
- Can be computed in $O(mn)$ using breadth first search, where m denotes # of edges.

Mean Geodesic distance of undirected networks

Definition

$$L = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij},$$

where n denotes # of nodes, and d_{ij} is the shortest distance between nodes i and j.

- Mean Geodesic distance includes distance to itself.
- Can be computed in $O(mn)$ using breadth first search, where m denotes # of edges.
- What happens if the network has multiple connected components?

Mean Geodesic distance of undirected networks

Definition

$$L = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij},$$

where n denotes # of nodes, and d_{ij} is the shortest distance between nodes i and j.

- Mean Geodesic distance includes distance to itself.
- Can be computed in $O(mn)$ using breadth first search, where m denotes # of edges.
- What happens if the network has multiple connected components?
- Harmonic mean (can have multiple connected components):

$$L^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}^{-1}$$

Summarization

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW `nd.edu`	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

Outline

Adjacency matrix

Definition

Given a finite graph $G = (V, E)$, an adjacency matrix A is a $|V| \times |V|$ matrix, whose elements indicate whether pairs of vertices are adjacent or not in the graph.

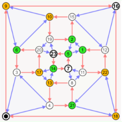
- The adjacency matrix is a (0,1)-matrix with zeros on its diagonal.
- If the graph is undirected, the adjacency matrix is symmetric.

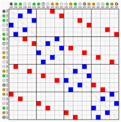
Adjacency matrix

Definition

Given a finite graph $G = (V, E)$, an adjacency matrix A is a $|V| \times |V|$ matrix, whose elements indicate whether pairs of vertices are adjacent or not in the graph.

- The adjacency matrix is a (0,1)-matrix with zeros on its diagonal.
- If the graph is undirected, the adjacency matrix is symmetric.



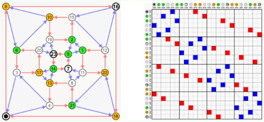


Adjacency matrix

Definition

Given a finite graph $G = (V, E)$, an adjacency matrix A is a $|V| \times |V|$ matrix, whose elements indicate whether pairs of vertices are adjacent or not in the graph.

- The adjacency matrix is a (0,1)-matrix with zeros on its diagonal.
- If the graph is undirected, the adjacency matrix is symmetric.



The adjacency matrix A of a bipartite graph whose two parts have r and s vertices can be written in the form
$$A = \begin{pmatrix} 0_{r,r} & B \\ B & 0_{s,s} \end{pmatrix}.$$

Storing a graph

Adjacency lists

An adjacency list is a collection of unordered lists used to represent a graph G. Each list describes the set of neighbors of a vertex in the graph.

1	2
1	11
1	10
2	3
3	4
4	5
5	6
5	8
5	11

(a)

1	4		
0	2		
1	3	5	
2	4	5	
0	3		
2	3	6	8
5	7		
6	8		
5	7		

(b)

1	2	-1
3	4	1
5	6	-1
7	8	-1
7	9	-1
10	11	1
12	13	1
14	15	1
16	17	1

(c)

Random walk of a graph

Markov chain

Suppose that $G = (V, E)$ is a graph of n vertices with vertex set V and edge set $E \subset V \times V$. Let $N(x) = \{y|(x, y) \in E\}$, and degree of vertex x denote as $d(x) = |N(x)|$.

Random walk of a graph

Markov chain

Suppose that $G = (V, E)$ is a graph of n vertices with vertex set V and edge set $E \subset V \times V$. Let $N(x) = \{y|(x, y) \in E\}$, and degree of vertex x denote as $d(x) = |N(x)|$.

- Note that x is isolated vertex if $N(x) = 0$.
- G is an undirected graph, we have $(x, y) \in E$ if $(y, x) \in E$.

Random walk of a graph

Markov chain

Suppose that $G = (V, E)$ is a graph of n vertices with vertex set V and edge set $E \subset V \times V$. Let $N(x) = \{y|(x, y) \in E\}$, and degree of vertex x denote as $d(x) = |N(x)|$.

- Note that x is isolated vertex if $N(x) = 0$.
- G is an undirected graph, we have $(x, y) \in E$ if $(y, x) \in E$.

For each $x \in V$, the transition matrix $P(y|x)$ is $\frac{1}{d(x)}$ if $y \in N(x)$, and $P(y|x) = 0$ otherwise.

Random walk of a graph

Markov chain

Suppose that $G = (V, E)$ is a graph of n vertices with vertex set V and edge set $E \subset V \times V$. Let $N(x) = \{y|(x, y) \in E\}$, and degree of vertex x denote as $d(x) = |N(x)|$.

- Note that x is isolated vertex if $N(x) = 0$.
- G is an undirected graph, we have $(x, y) \in E$ if $(y, x) \in E$.

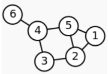
For each $x \in V$, the transition matrix $P(y|x)$ is $\frac{1}{d(x)}$ if $y \in N(x)$, and $P(y|x) = 0$ otherwise.

- Let X be a random walk on G, if G is connected then X is irreducible.
- X has period 2 if and only if G is bipartite, in which case the parts are the cyclic classes of X.
- Let $D = diag(d_1, d_2, \cdots, d_n)$ be a diagonal matrix, and $P = D^{-1}A$.

Combinatorial Laplacian of graph

Definition

Labeled graph



Degree matrix

$$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Adjacency matrix

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

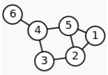
Laplacian matrix

$$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

Combinatorial Laplacian of graph

Definition

Labeled graph



Degree matrix

$$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Adjacency matrix

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

Laplacian matrix

$$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

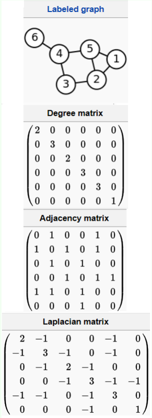
Given a graph G, (Combinatorial) Laplacian of G: $L = D - A$, i.e.,

$$L(u, v) = \begin{cases} d_v, & \text{if } u = v; \\ -1, & \text{if } u \text{ and } v \text{ are adjacent ;} \\ 0, & \text{otherwise.} \end{cases}$$

If G is an undirected graph G, and its Laplacian matrix L with eigenvalues $\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1}$, then

Combinatorial Laplacian of graph

Definition



Given a graph G, (Combinatorial) Laplacian of G: $L = D - A$, i.e.,

$$L(u,v) = \begin{cases} d_v, & \text{if } u = v; \\ -1, & \text{if } u \text{ and } v \text{ are adjacent ;} \\ 0, & \text{otherwise.} \end{cases}$$

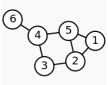
If G is an undirected graph G, and its Laplacian matrix L with eigenvalues $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$, then

- L is singular and symmetric(existing $\lambda_i = 0$).

Combinatorial Laplacian of graph

Definition

Labeled graph



Degree matrix

$$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Adjacency matrix

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

Laplacian matrix

$$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

Given a graph G, (Combinatorial) Laplacian of G:
$L = D - A$, i.e.,

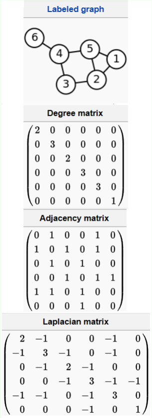
$$L(u,v) = \begin{cases} d_v, & \text{if } u = v; \\ -1, & \text{if } u \text{ and } v \text{ are adjacent }; \\ 0, & \text{otherwise.} \end{cases}$$

If G is an undirected graph G, and its Laplacian matrix L with eigenvalues $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$, then

- L is singular and symmetric(existing $\lambda_i = 0$).
- Since row sum and column sum of L is zero, $\lambda_0 = 0$ and $v_0 = (1, 1, \cdots, 1)$.

Combinatorial Laplacian of graph

Definition



Given a graph G, (Combinatorial) Laplacian of G: $L = D - A$, i.e.,

$$L(u,v) = \begin{cases} d_v, & \text{if } u = v; \\ -1, & \text{if } u \text{ and } v \text{ are adjacent }; \\ 0, & \text{otherwise.} \end{cases}$$

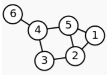
If G is an undirected graph G, and its Laplacian matrix L with eigenvalues $\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1}$, then

- L is singular and symmetric(existing $\lambda_i = 0$).
- Since row sum and column sum of L is zero, $\lambda_0 = 0$ and $v_0 = (1, 1, \cdots, 1)$.
- The second smallest eigenvalue is called algebraic connectivity.

Combinatorial Laplacian of graph

Definition

Labeled graph



Degree matrix

$$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Adjacency matrix

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

Laplacian matrix

$$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

Given a graph G, (Combinatorial) Laplacian of G: $L = D - A$, i.e.,

$$L(u,v) = \begin{cases} d_v, & \text{if } u = v; \\ -1, & \text{if } u \text{ and } v \text{ are adjacent ;} \\ 0, & \text{otherwise.} \end{cases}$$

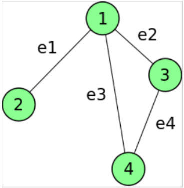
If G is an undirected graph G, and its Laplacian matrix L with eigenvalues $\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1}$, then

- L is singular and symmetric(existing $\lambda_i = 0$).
- Since row sum and column sum of L is zero, $\lambda_0 = 0$ and $v_0 = (1, 1, \cdots, 1)$.
- The second smallest eigenvalue is called algebraic connectivity.
- For weighted graph G, Laplacian can be defined in a same manner.

Incidence matrix

Definition

An incidence matrix B is a $|V| \times |E|$ matrix that shows the relationship between vertices and edges of graph $G = (V, E)$.



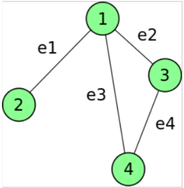
$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & -1 \end{pmatrix}$$

Incidence matrix

Definition

An incidence matrix B is a $|V| \times |E|$ matrix that shows the relationship between vertices and edges of graph $G = (V, E)$.

- Each column corresponds to an edge $e = (v_i, v_j)$ (with $i < j$), where the value of an entry is 1 in the row corresponding to v_i, and entry -1 in the row corresponding to v_j.



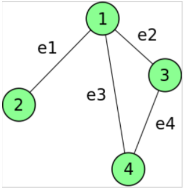
$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & -1 \end{pmatrix}$$

Incidence matrix

Definition

An incidence matrix B is a $|V| \times |E|$ matrix that shows the relationship between vertices and edges of graph $G = (V, E)$.

- Each column corresponds to an edge $e = (v_i, v_j)$ (with $i < j$), where the value of an entry is 1 in the row corresponding to v_i, and entry -1 in the row corresponding to v_j.
- $L = BB^T$. Thus, L is positive semidefinite and has nonnegative eigenvalues since $\mathbf{x}^T L\mathbf{x} = \mathbf{x}^T BB^T\mathbf{x} = (B^T\mathbf{x})^T(B^T\mathbf{x}) \geq 0$ $(\lambda_i \geq 0)$.



$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & -1 \end{pmatrix}$$

Normalized Laplacian of graph

Definition

Given a graph G, normailized Laplacian of G: $\mathcal{L} = D^{-1/2}LD^{-1/2}$,

i.e., $\mathcal{L}(u,v) = \begin{cases} 1, & \text{if } u = v; \\ -\frac{1}{\sqrt{d_u d_v}}, & \text{if } u \text{ and } v \text{ are adjacent ;} \\ 0, & \text{otherwise.} \end{cases}$

Normalized Laplacian of graph

Definition

Given a graph G, normailized Laplacian of G: $\mathcal{L} = D^{-1/2}LD^{-1/2}$,

i.e., $\mathcal{L}(u,v) = \begin{cases} 1, & \text{if } u = v; \\ -\frac{1}{\sqrt{d_u d_v}}, & \text{if } u \text{ and } v \text{ are adjacent }; \\ 0, & \text{otherwise.} \end{cases}$

- $\mathcal{L} = D^{-1/2}BB^T D^{-1/2} = I - D^{-1/2}AD^{-1/2} = D^{1/2}(I-P)D^{-1/2}$. Thus, $\mathcal{L}$ is positive semidefinite and $0 \leq \lambda(\mathcal{L}) \leq 2$.

Normalized Laplacian of graph

Definition

Given a graph G, normailized Laplacian of G: $\mathcal{L} = D^{-1/2}LD^{-1/2}$,

i.e., $\mathcal{L}(u,v) = \begin{cases} 1, & \text{if } u = v; \\ -\frac{1}{\sqrt{d_u d_v}}, & \text{if } u \text{ and } v \text{ are adjacent }; \\ 0, & \text{otherwise.} \end{cases}$

- $\mathcal{L} = D^{-1/2}BB^TD^{-1/2} = I - D^{-1/2}AD^{-1/2} = D^{1/2}(I-P)D^{-1/2}$. Thus, $\mathcal{L}$ is positive semidefinite and $0 \le \lambda(\mathcal{L}) \le 2$.
- L is singular and symmetric, and $\lambda_0 = 0$ corresponding to eigenvector $D^{1/2}v_0^T = D^{1/2}(1,1,\cdots,1)^T$.

Normalized Laplacian of graph

Definition

Given a graph G, normailized Laplacian of G: $\mathcal{L} = D^{-1/2}LD^{-1/2}$, i.e., $\mathcal{L}(u,v) = \begin{cases} 1, & \text{if } u = v; \\ -\frac{1}{\sqrt{d_u d_v}}, & \text{if } u \text{ and } v \text{ are adjacent ;} \\ 0, & \text{otherwise.} \end{cases}$

- $\mathcal{L} = D^{-1/2}BB^TD^{-1/2} = I - D^{-1/2}AD^{-1/2} = D^{1/2}(I-P)D^{-1/2}$. Thus, $\mathcal{L}$ is positive semidefinite and $0 \le \lambda(\mathcal{L}) \le 2$.
- L is singular and symmetric, and $\lambda_0 = 0$ corresponding to eigenvector $D^{1/2}v_0^T = D^{1/2}(1,1,\cdots,1)^T$.
- P has an eigenvalue $1-\lambda_i$, where λ_i is an eigenvalue of $\mathcal{L}$.

Normalized Laplacian of graph

Definition

Given a graph G, normailized Laplacian of G: $\mathcal{L} = D^{-1/2}LD^{-1/2}$,

i.e., $\mathcal{L}(u,v) = \begin{cases} 1, & \text{if } u = v; \\ -\frac{1}{\sqrt{d_u d_v}}, & \text{if } u \text{ and } v \text{ are adjacent ;} \\ 0, & \text{otherwise.} \end{cases}$

- $\mathcal{L} = D^{-1/2}BB^TD^{-1/2} = I - D^{-1/2}AD^{-1/2} = D^{1/2}(I-P)D^{-1/2}$. Thus, $\mathcal{L}$ is positive semidefinite and $0 \leq \lambda(\mathcal{L}) \leq 2$.
- L is singular and symmetric, and $\lambda_0 = 0$ corresponding to eigenvector $D^{1/2}v_0^T = D^{1/2}(1, 1, \cdots, 1)^T$.
- P has an eigenvalue $1 - \lambda_i$, where λ_i is an eigenvalue of $\mathcal{L}$.
- The regularization of graph G: $F^T\mathcal{L}F = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}}\right)^2$.

Properties of normalized Laplacian [WebScience 2013]

Properties

The eigenvalues of the normalized Laplacian matrix of graph G with n vertices satisfy the following properties:

- $0 \leq \lambda_2 \leq \frac{n}{n-1} \leq \lambda_n \leq 2$.
- $\lambda_2 = \cdots = \lambda_n = \frac{n}{n-1}$ if and only if G is a clique.
- $\lambda_n = 2$ if and only if G is a bi-clique.
- G has at least i connected components if and only if $\lambda_j = 0$, for $j = 1, 2, \cdots, i$.
- The mean of eigenvalues $\lambda_2, \lambda_3, \cdots, \lambda_n$ of a network G with n vertices is $\frac{n}{n-1}$.
- The variance of eigenvalues $\lambda_2, \lambda_3, \cdots, \lambda_n$ of a network G with n vertices is $\frac{1}{n-1} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{A_{ij}}{d(v_i)d(v_j)} - \frac{n}{(n-1)^2}$ (R-energy).

Network generation

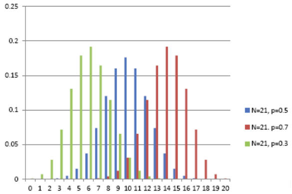
Generators

- Erdös-Renyi model
- Preferential attachment
- Variations + extensions
  - Copying model
  - Triad-closing
  - Butterfly model
- Recursion - Kronecker generator

Random network generator: Erdös-Renyi model

Erdös-Renyi model is known as the random graph model, which generates undirected random graphs.

- Parameters: N (# vertices) and p (prob. of forming an edge)
- For each possible node pair, the approach generates an edge with probability p. Thus, # edges $= \frac{pN(N-1)}{2}$.
- Degree distribution:
  - $P(\text{node has degree k}) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$
  - Follows binomial distribution with mean $(N-1)p$ and variance $(N-1)p(1-p)$ (not power-law distribution).



Scale-free network generator

Preferential attachment model

The more connected a node is, the more likely it is to receive new links (namely, Rich gets Richer, Matthew Effect or Paretos Law, etc.).

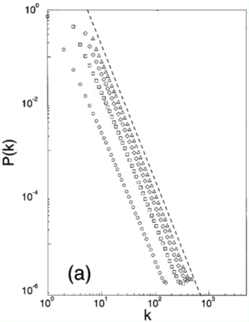
- Price model
- Barabasi Albert model

Price model for citation networks

- Each new paper is generated with m citations (mean).
- New papers cite previous papers with probability proportional to their indegree (citations).
  - Each new paper is generated with m citations (mean).
  - New papers cite previous papers with probability proportional to their indegree (citations).
  - Power law with exponent $\alpha = 2 + \frac{1}{m}$ [Science 1965]

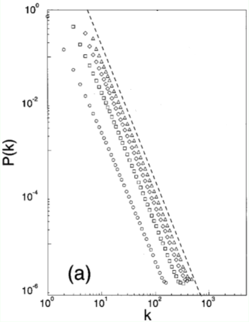
Barabasi Albert model

Model



Barabasi Albert model

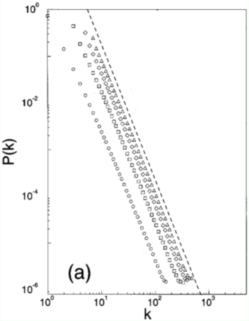
Model



- Start with an initial network of m_0 (≥ 2) nodes, and the degree of each node ≥ 1, otherwise it will always remain isolated.

Barabasi Albert model

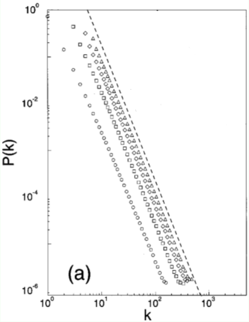
Model



- Start with an initial network of m_0 (≥ 2) nodes, and the degree of each node ≥ 1, otherwise it will always remain isolated.
- For each new node, connect it to m existing nodes i with a probability p_i, where $p_i = \frac{k_i}{\sum_j k_j}$, where k_i is degree of node i.

Barabasi Albert model

Model



- Start with an initial network of m_0 (≥ 2) nodes, and the degree of each node ≥ 1, otherwise it will always remain isolated.
- For each new node, connect it to m existing nodes i with a probability p_i, where $p_i = \frac{k_i}{\sum_j k_j}$, where k_i is degree of node i.
- Results in a single connected component with power-law degree distribution with $\alpha = 3$ [Reviews of Modern Physics 2003].

Kronecker product of matrices

Given two matrices $U \in \mathbb{R}^{n \times m}$ and $V \in \mathbb{R}^{p \times q}$, the Kronecker product matrix $S \in \mathbb{R}^{np \times mq}$ is given by

$$S = U \bigotimes V = \begin{pmatrix} u_{11}V & u_{12}V & \cdots & u_{1m}V \\ u_{21}V & u_{22}V & \cdots & u_{2m}V \\ \cdots & \cdots & \cdots & \cdots \\ u_{n1}V & u_{n2}V & \cdots & u_{nm}V \end{pmatrix}$$

Kronecker product of matrices

Given two matrices $U \in \mathbb{R}^{n \times m}$ and $V \in \mathbb{R}^{p \times q}$, the Kronecker product matrix $S \in \mathbb{R}^{np \times mq}$ is given by

$$S = U \bigotimes V = \begin{pmatrix} u_{11}V & u_{12}V & \cdots & u_{1m}V \\ u_{21}V & u_{22}V & \cdots & u_{2m}V \\ \cdots & \cdots & \cdots & \cdots \\ u_{n1}V & u_{n2}V & \cdots & u_{nm}V \end{pmatrix}$$

- $A \bigotimes (aB + C) = (aA) \bigotimes B + A \bigotimes C$, but $A \bigotimes B \neq B \bigotimes A$.

Kronecker product of matrices

Given two matrices $U \in \mathbb{R}^{n \times m}$ and $V \in \mathbb{R}^{p \times q}$, the Kronecker product matrix $S \in \mathbb{R}^{np \times mq}$ is given by

$$S = U \bigotimes V = \begin{pmatrix} u_{11}V & u_{12}V & \cdots & u_{1m}V \\ u_{21}V & u_{22}V & \cdots & u_{2m}V \\ \cdots & \cdots & \cdots & \cdots \\ u_{n1}V & u_{n2}V & \cdots & u_{nm}V \end{pmatrix}$$

- $A \bigotimes (aB + C) = (aA) \bigotimes B + A \bigotimes C$, but $A \bigotimes B \neq B \bigotimes A$.
- $(A \bigotimes B)(C \bigotimes D) = (AC) \bigotimes (BD)$.

Kronecker product of matrices

Given two matrices $U \in \mathbb{R}^{n \times m}$ and $V \in \mathbb{R}^{p \times q}$, the Kronecker product matrix $S \in \mathbb{R}^{np \times mq}$ is given by

$$S = U \bigotimes V = \begin{pmatrix} u_{11}V & u_{12}V & \cdots & u_{1m}V \\ u_{21}V & u_{22}V & \cdots & u_{2m}V \\ \cdots & \cdots & \cdots & \cdots \\ u_{n1}V & u_{n2}V & \cdots & u_{nm}V \end{pmatrix}$$

- $A \bigotimes (aB + C) = (aA) \bigotimes B + A \bigotimes C$, but $A \bigotimes B \neq B \bigotimes A$.
- $(A \bigotimes B)(C \bigotimes D) = (AC) \bigotimes (BD)$.
- $(A \bigotimes B)^{-1} = A^{-1} \bigotimes B^{-1}$ and $(A \bigotimes B)^T = A^T \bigotimes B^T$

Kronecker product of matrices

Given two matrices $U \in \mathbb{R}^{n \times m}$ and $V \in \mathbb{R}^{p \times q}$, the Kronecker product matrix $S \in \mathbb{R}^{np \times mq}$ is given by

$$S = U \bigotimes V = \begin{pmatrix} u_{11}V & u_{12}V & \cdots & u_{1m}V \\ u_{21}V & u_{22}V & \cdots & u_{2m}V \\ \cdots & \cdots & \cdots & \cdots \\ u_{n1}V & u_{n2}V & \cdots & u_{nm}V \end{pmatrix}$$

- $A \bigotimes (aB + C) = (aA) \bigotimes B + A \bigotimes C$, but $A \bigotimes B \neq B \bigotimes A$.
- $(A \bigotimes B)(C \bigotimes D) = (AC) \bigotimes (BD)$.
- $(A \bigotimes B)^{-1} = A^{-1} \bigotimes B^{-1}$ and $(A \bigotimes B)^T = A^T \bigotimes B^T$
- $|A \bigotimes B| = |A|^m |B|^n$ and $Tr(A \bigotimes B) = Tr(A) Tr(B)$ if $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$.

Kronecker product of matrices

Given two matrices $U \in \mathbb{R}^{n\times m}$ and $V \in \mathbb{R}^{p\times q}$, the Kronecker product matrix $S \in \mathbb{R}^{np\times mq}$ is given by

$$S = U \bigotimes V = \begin{pmatrix} u_{11}V & u_{12}V & \cdots & u_{1m}V \\ u_{21}V & u_{22}V & \cdots & u_{2m}V \\ \cdots & \cdots & \cdots & \cdots \\ u_{n1}V & u_{n2}V & \cdots & u_{nm}V \end{pmatrix}$$

- $A \bigotimes (aB + C) = (aA) \bigotimes B + A \bigotimes C$, but $A \bigotimes B \neq B \bigotimes A$.
- $(A \bigotimes B)(C \bigotimes D) = (AC) \bigotimes (BD)$.
- $(A \bigotimes B)^{-1} = A^{-1} \bigotimes B^{-1}$ and $(A \bigotimes B)^T = A^T \bigotimes B^T$
- $|A \bigotimes B| = |A|^m |B|^n$ and $Tr(A \bigotimes B) = Tr(A)Tr(B)$ if $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times m}$.
- We define $k-$th Kronecker power of A_1 as $A_1^{[k]}$ (abbreviated to A_k), where $A_k = A_1^{[k]} = A_{k-1} \bigotimes A_1$.

Kronecker model cont.

Model

Instead of a single property of the network, Kronecker model can fit multiple properties of a network, which makes them interesting for fitting.

Kronecker model cont.

Model

Instead of a single property of the network, Kronecker model can fit multiple properties of a network, which makes them interesting for fitting.

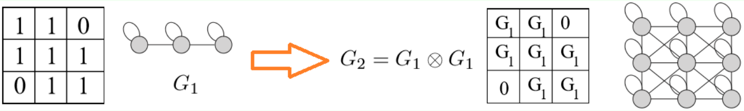
- Deterministic Kronecker model: it begins with an initiator graph G_1 with N_1 nodes, and produces successively larger graphs $G_2, \cdots, G_n$ such that the $k-$th graph G_k has $N_k = N_1^k$.

Kronecker model cont.

Model

Instead of a single property of the network, Kronecker model can fit multiple properties of a network, which makes them interesting for fitting.

- Deterministic Kronecker model: it begins with an initiator graph G_1 with N_1 nodes, and produces successively larger graphs $G_2, \cdots, G_n$ such that the $k-$th graph G_k has $N_k = N_1^k$.
- Stochastic Kronecker model: it starts with a $N_1 \times N_1$ probability matrix $\Theta = [\theta_{ij}]$, where the element $\theta_{ij} \in [0, 1]$ is the probability that edge (i, j) is present.



Sources for generator

Generators

- Erdös Renyi: http://ladamic.com/netlearn/NetLogo501/ErdosRenyiDegDist.html
- BRITE: http://wwwcsbuedu/brite/
- INET: http://topology.eecs.umich.edu/inet
- Kronecker:
  - christos@cs.cmu.edu
  - http://www.cc.gatech.edu/dimacs10/archive/kronecker.shtml
  - http://www.cc.gatech.edu/dimacs10/archive/kronecker.shtml

Take-home messages

- Graph
  - Motivations
  - Patterns
- Graph aspects
  - Graph types
  - Properties
  - Graph modeling
- Network generation
  - Erdös Renyi model
  - Barabasi Albert model
  - Kronecker model