Social Network Analysis: Descriptives Part 2

EPIC - SNA, Columbia University

Zack W Almquist

June 12th, 2018

University of Minnesota

Degree Distribution

Network Composition

Network Position (Theories)

Distance Measures

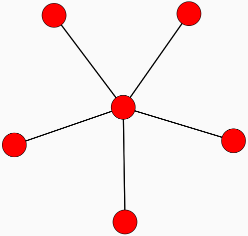
R for Descriptive Analysis: More tools

Degree Distribution

Degree Distribution: Local Level

Given some ego, what do we want to know about their alters?

- Network Size
  - Degree - total number of ties incident to a node
    - E.g., GSS "Important matters" networks: 2.9(1985);2(2004)
    - E.g, "Dunbar Number": 150-200



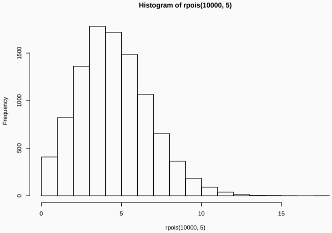
Degree Estimation

Not just central tendencies, but distribution(s) too.

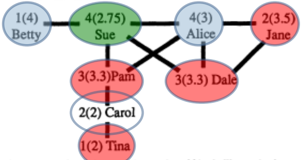
- How many friends do most people have?

vs.

- How many friends "should" I have?



Degree Estimation



The number beside each name is her number of friends. The number in parentheses beside each name is the mean number of friends of her frier

FIG. 1.—Friendships among eight girls at Marketville

Betty: 4/1 = 4

Alice: (4 + 3 + 3 + 3) / 4 = 3

TABLE 1

A SUMMARY OF THE NUMBERS OF FRIENDS AND THE MEAN NUMBERS OF FRIENDS OF FRIENDS FOR EACH OF THE GIRLS IN FIGURE 1

	Number of Friends (x_i)	Total Number of Friends of Her Friends (Σx_j)	Mean Number of Friends of Her Friends ($\Sigma x_j/x_i$)
Betty	1	4	4
Sue	4	11	2.75
Alice	4	12	3
Jane	2	7	3.5
Pam	3	10	3.3
Dale	3	10	3.3
Carol	2	4	2
Tina	1	2	2
Total	20	60	23.92
Mean	2.5*	3†	2.99*

* For eight girls.
† For 20 friends.

jimi adams. EPIC- SNA 2017. Columbia University. Origin: Feld SL. Why Your Friends have More Friends than you do. American Journal of Sociology 1991;96:1464-1477.

Degree Estimation

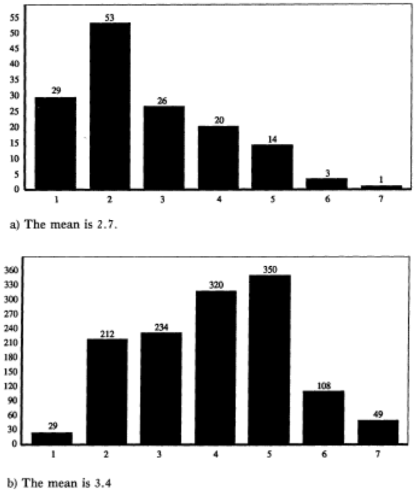


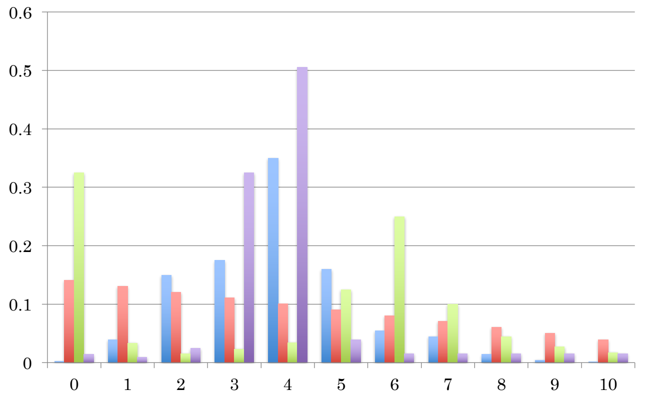
FIG. 3.—(*a*) Distribution of numbers of friends for Marketville girls; (*b*) distribution of number of friends' friends for Marketville girls.

Friends
mean = 2.6
median = 2

Friends' friends
mean = 3.97
median = 4

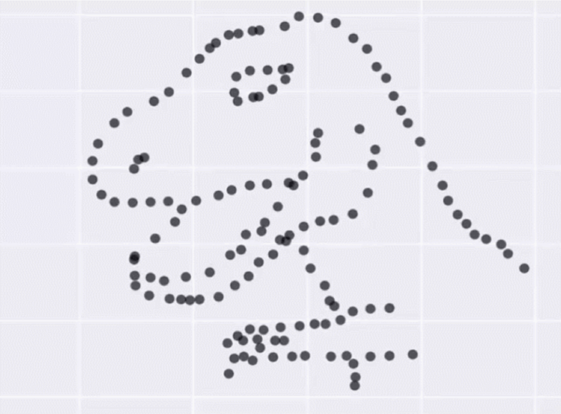
jimi adams, ERIC, SNA 2017, Columbia University. Origin: Feld SL. Why Your Friends have More Friends than

Degree Distributions Matter



jimi adams. EPIC- SNA 2017. Columbia University. Origin:

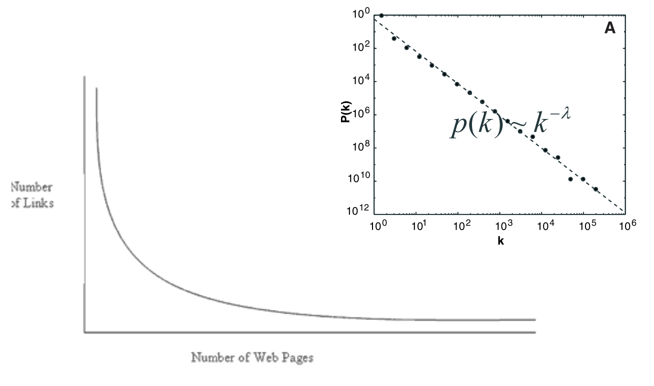
Degree Distributions Matter



Each dataset has the same summary statistics to two decimal places: ($\bar{x}$ =54.26, $\bar{y}$ = 47.83, sd_x = 16.76, sd_y = 26.93, Pearson's r = -0.06). https://www.autodeskresearch.com/publications/samestats

jimi adams. EPIC- SNA 2017. Columbia University. Origin:

Degree Distributions Matter

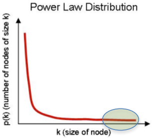


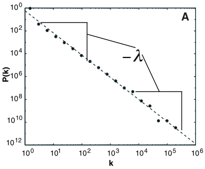
jimi adams. EPIC- SNA 2017. Columbia University. Origin: Barabasi AL, Albert R, Jeong H, Bianconi G. Power-Law Distribution of the World Wide Web. Science 2000;287:2115.

Degree Distributions Matter

- A scale-free network has two basic properties:
  - Expands continuously by adding new vertices
  - New vertices attach preferentially to vertices that are already well connected
    - at a constant rate

∴ "Hubs" seen as important





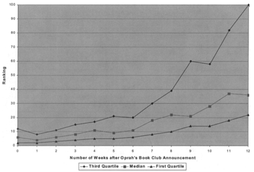
jimi adams. EPIC- SNA 2017. Columbia University. Origin: Barabasi AL, Albert R, Jeong H, Bianconi G. Power-Law Distribution of the World Wide Web. Science 2000;287:2115.

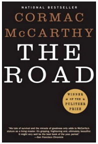
Degree Distributions Matter

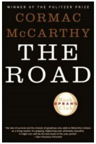
- Oprah Book Club
  - Before endorsement 5/45 in top 150
  - The week after:
    - All were
    - Most remained for months
    - Even for re-issues
    - Increased duration

Figure 1

Oprah's Picks' Rankings Following Club Induction



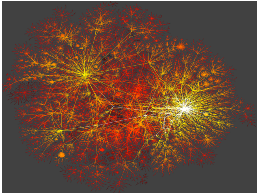


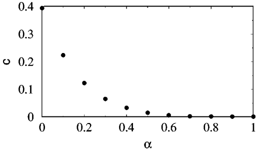


jimi adams. EPIC- SNA 2017. Columbia University. Origin:Butler RJ, Cowan BW, Nilsson S. From Obscurity to Bestseller: Examining the Impact of Oprah's Book Club Selections. Publishing Research Quarterly 2005;20(4):23-34

Degree Distributions Matter

- Given a scale-free network topology, hubs can facilitate inoculation strategies:
  - w/ Increased success targeting hubs (α)…
  - need reduced # of applied cures (c)…
  - to achieve crossing the same "epidemic threshold."

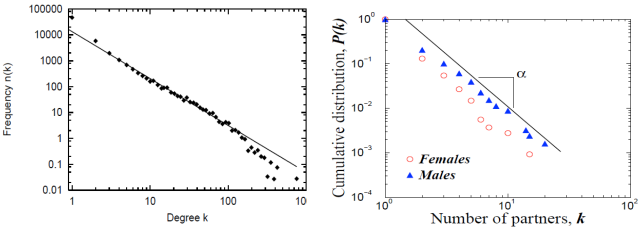




jimi adams. EPIC- SNA 2017. Columbia University. Origin: Dezso Z, Barabasi AL. Halting viruses in scale-free Networks. Physical Review E 2002.

Degree Distributions Matter

Can we move from stopping viruses on the WWW to preventing STDs?



jimi adams. EPIC- SNA 2017. Columbia University. Origin: Dezso Z, Barabasi AL. Halting viruses in scale-free Networks. Physical Review E 2002.

Degree Distributions Matter

- Scale-free networks are typically identified by degree-curve fitting.
- And are *assumed* to follow preferential attachment processes

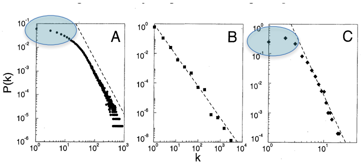
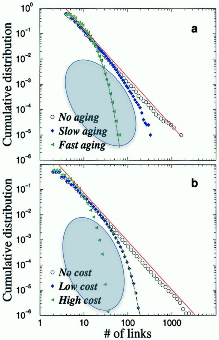


Fig. 1. The distribution function of connectivities for various large networks. **(A)** Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. **(B)** WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). **(C)** Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.



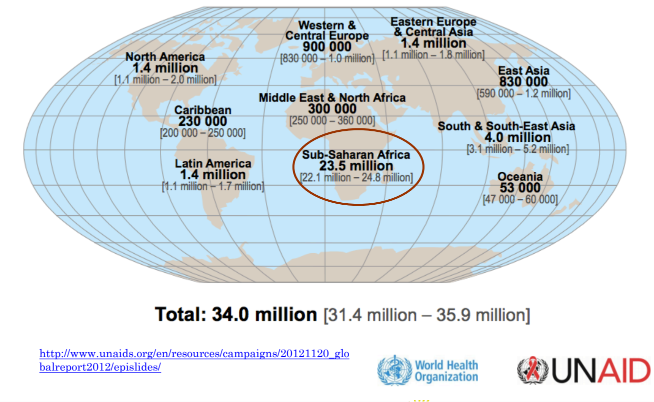
jimi adams. EPIC- SNA 2017. Columbia University. Origin: Amaral LAN, Scala A, Barthelemy M, Stanley HE. Classes of Small World Networks. PNAS 2000;97(21):1

Degree Distributions Matter

- Theoretically, any degree-based metric has ***no necessary*** relation to a network's structure.
  - There are many graphs with *identical* degree distributions but very different topologies (enter ergm).
- Preferential attachment → scale free
  - Not (necessarily) vice versa
- Even finding a power-law degree distribution (but see previous slide), isn't enough to say much about global network structure.
  - (e.g., If *any* non-degree based patterns underlie network formation processes.)

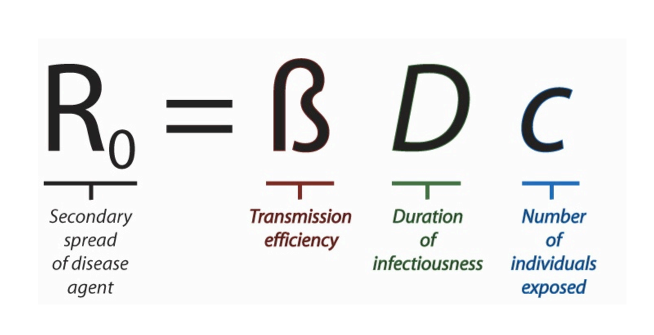
Degree Distributions Matter

Adults & Children Estimated to be Living w/HIV, 2011



jimi adams. EPIC- SNA 2017. Columbia University. Origin:

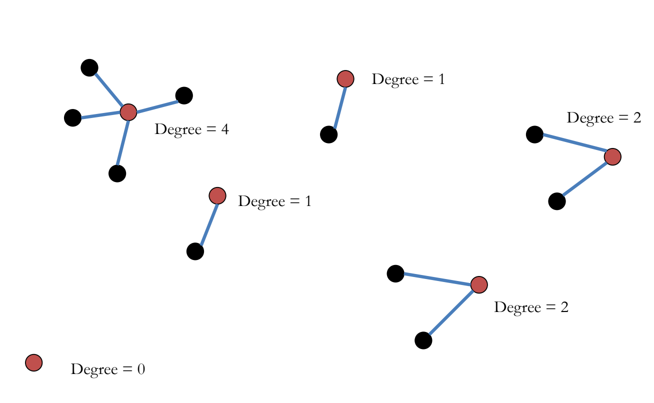
Degree Distributions Matter



jimi adams. EPIC- SNA 2017. Columbia University. Origin:May RM, Anderson RM. Transmission Dynamics of HIV Infection. Nature 1987;326:137-142.

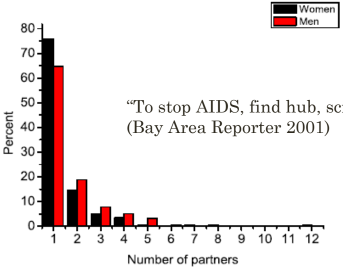
Degree Distributions Matter

Who is at greatest/least risk?



jimi adams. EPIC- SNA 2017. Columbia University. Origin:

STI Interventions



"To stop AIDS, find hub, scientists say."
(Bay Area Reporter 2001)

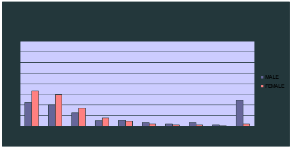
Figure 1: Distribution of the number of sex partners during one year for men and women in the Gotland data.

Degree Distributions Matter

Uganda 18% HIV+

1994

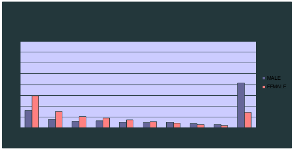
(*Rakai Sexnet* study)



United States 1% HIV+

1994

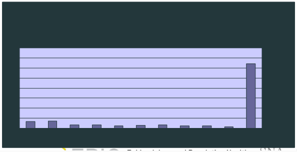
(*NHSLS* study)



Thailand 2% HIV+

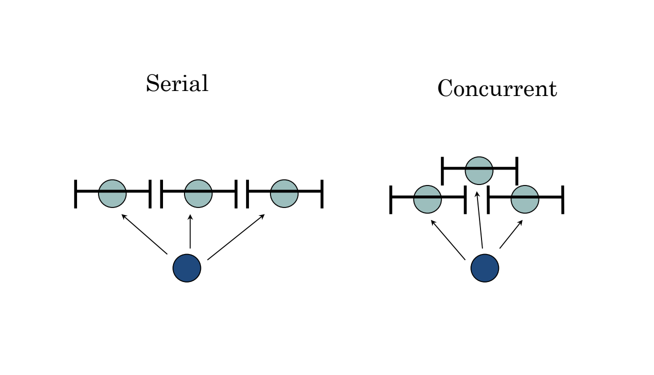
1993

(*BRAIDS* study)



jimi adams. EPIC- SNA 2017. Columbia University. Origin: Liljeros F, Edling CR, Amaral LAN, Stanley HE, Aberg Y. The Web of Human Sexual Contacts. Nature 2001;411:9

Degree Distributions Matter



jimi adams. EPIC- SNA 2017. Columbia University. Origin: Morris (2003)

Degree Distributions Matter

	Uganda	US	Thailand
Total any concurrency	27.1	14.3	29.2
Pair Composition			
2 long term	92.2	88.0	35.9
1 long, 1 short	7.8	12.0	63.8

Thailand has higher levels of short-term paired with long-term

Observation	Uganda	US	Thailand
Completed	10.0 (59%)	4.0 (78%)	0.03 (87%)
Ongoing	36.0 (41%)	9.5 (22%)	23.5 (13%)

Median Overlap (months)

So the duration of overlap is much shorter

Source: Morris (2003)

jimi adams. EPIC- SNA 2017. Columbia University. Origin: Morris (2003)

Degree Distributions Matter

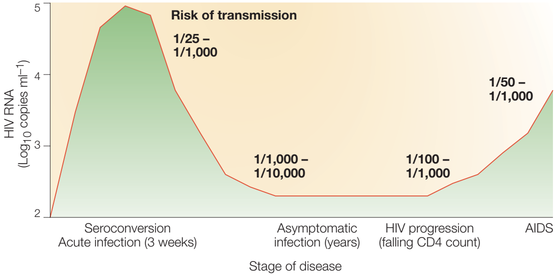
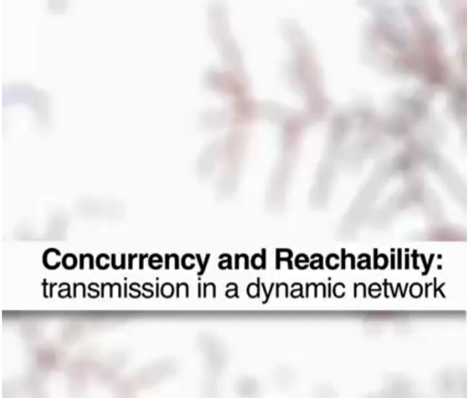


Figure 1 | **The changing viral load during the different stages of disease and the effects of viral load on the probability of sexual transmission of HIV.**

jimi adams. EPIC- SNA 2017. Columbia University. Origin: Galvin SR, COhen MS. The Role of Sexually Transmitted Diseases in HIV Transmission. Nature Reviews: Microbiology 2004;2:33-42.

Degree Distributions Matter



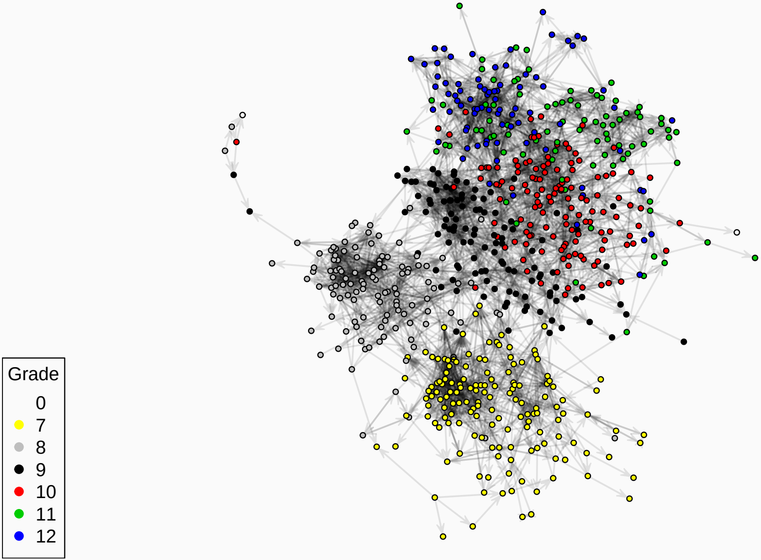
Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S.

Concurrent Partnerships and HIV Prevalence Disparities by Race: Linking Science and Public Health Practice.

American Journal of Public Health 2009;99(6):1023-1031.

Network Composition

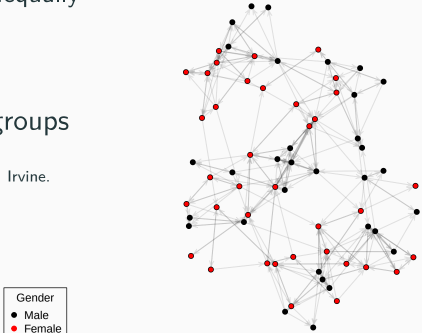
Homophily: Add Health Friendship Network



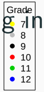
Selective Mixing

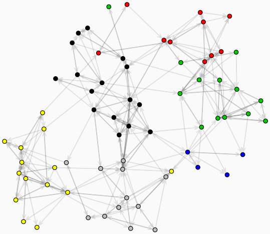
- Observation: vertices in most real-world networks are heterogeneous
  - Different classes of vertices unequally likely to be tied
- Selective mixing: differences in marginal tie probability among groups

Carter Butts. Social Network Methods. University of California, Irvine.



- Very common type of selective mixing: marginal tendency of individuals with similar properties to be adjacent
  - Term refers to association, not underlying mechanism
- Can occur for many reasons
  - Active selection (i.e., preferences)
  - Passive selection (e.g., tracking in schools)
  - Diffusion/influence
- Less well-understood than one would think



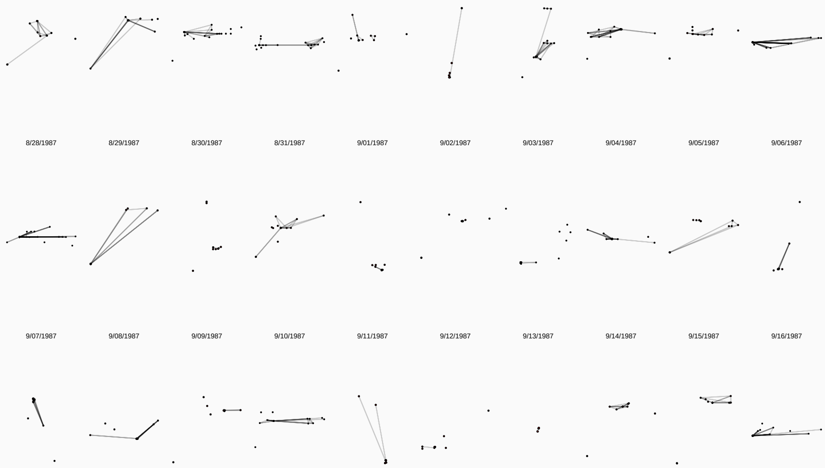


Carter Butts. Social Network Methods. University of California, Irvine.

Propinquity and Age Mixing

- More complex mixing is also possible
- Propinquity: tendency of spatially proximate vertices to be tied
  - Continuous covariate – function form of effect important for resulting structure
- Age mixing: differential association based on age
  - Often broadly homophilous, but details can vary
  - Asymmetries common, e.g., male/female age difference in marriages

Freeman et al. (1988)



Mixing Rates: Some Basics

- Observation: we already know how to calculate one kind of mixing rate – the density!
  - Rate at which all vertices mix with all other vertices
- Generalization to multiple groups
  - Let S, $S^{'}$ be two groups. Then

$$r_{SS'} = \frac{\sum_{i \in S} \sum_{j \in S'} Y_{ij}}{|\{(s, s') : s \in S, s' \in S', s \neq s')\}|}$$

  - Not that this reduces to density when $S = S^{'} = V$ (loops omitted here)
- Relationship to inhomogeneous Bernoulli Model
  - For $i \in S$, $j \in S^{'}$ let $\Phi_{ij} = r_{SS'}$. Bernoulli model with parameter Φ expresses distribution in which edges are independent and homogeneous within groups, with fixed inter/intra-group mixing rates

Null Distributions

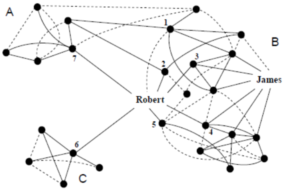
- Several ways to obtain null distributions
  - Binomial model
    - Let $X_{SS'}$ be the number of edges from S to $S^{'}$, and let $X^*_{SS'}$ be the number of possible edges. If all $S, S^{'}$ edge variables are independent with probability p, then $X_{SS'} \sim Binom(X^*_{SS'}, p)$; its expected value is thus $pX^*_{SS'}$ and its variance is $p(1-p)X^*_{SS'}$
    - Can get null from assumption that p is equal to the density
  - Poisson approximation
    - If $X^*_{SS'} \to \infty$ with $pX^*_{SS'}$ fixed, then $X^*_{SS'} \sim Pois(pX^*_{SS'})$; expected value and variance in this case are both $pX^*_{SS'}$
    - Can get null expectation using two-way marginals: $pX^*_{SS'} = X_S X_{S'} / (\sum_i \sum_j Y_{ij})$; comparison using z-scores typically works fairly well for exploratory purposes
  - Permutation of vertex labels
    - Randomly re-assign set memberships, holding set sizes constant, and recalculate $X_{SS'}$. This both exact, and preserves edge value distributions (useful if valued)

Network Position (Theories)

Structural Holes

Who's in an advantageous network position?

- James Coleman – social capital derives from social closure
- Ron Burt – social capital derives from bridging non-overlapping contacts



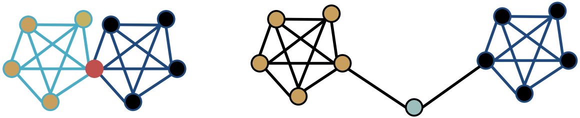
Density Table of Relations Within and Between Groups

.65			Group A (5 people and 8 ties; 5 strong, 3 weak)
.05	.25		Group B (17 people and 41 ties; 27 strong, 14 weak)
.00	.01	.65	Group C (5 people and 8 ties; 5 strong, 3 weak)

jimi adams. EPIC- SNA 2017. Columbia University. Origin: Burt RS. "Structural Holes versus Network Closure as Social Capital." Pp.31-56ininLinN,CookK,BurtRS.Social

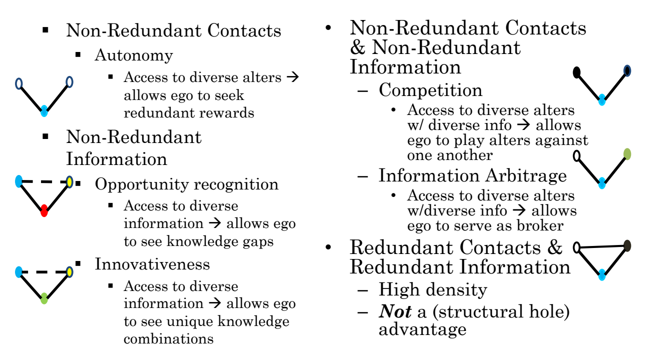
Structural Holes

- Structural Holes –
  - Gap between two sets of nodes
  - The structural hole advantage – people accrue capital by occupying positions with a maximum of non-redundant contacts
    - Attained by filling gaps (structural holes) between otherwise disconnected sub-graphs
- *Tertius Gaudens*
  - Literally – the third who benefits



jimi adams. EPIC- SNA 2017. Columbia University. Origin: Burt RS. Structural Holes: The social structure of competition. Harvard University Press; 1992.

Structural Holes



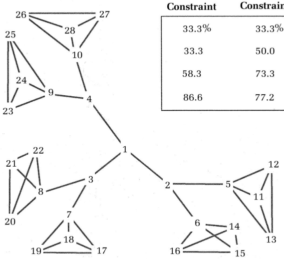
jimi adams. EPIC- SNA 2017. Columbia University. Origin: Rodan S. Structural Holes and Managerial Performance:Identifying Underlying Mechanisms. Social Networks.2010;32(3).

Structural Holes

FIGURE 2
Illustration of Direct and Indirect Network Constraint

$$C_{ij} = \left(p_{ij} + \sum_q p_{iq} p_{qj} \right)^2$$

Direct Network Constraint	Indirect Network Constraint	Role in Network
33.3%	33.3%	Broker of brokers (no.1)
33.3	50.0	Broker (nos. 2, 3, 4)
58.3	73.3	Group leader (nos. 5 to 10)
86.6	77.2	Group member (nos. 11 to 28)



"where p_{ij} is the proportion of i's network time and energy invested in contact j,

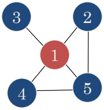
$p_{ij} = \Sigma z_{ij} / \Sigma_q z_{iq}$, $(q \neq i, j)$

and variable z_{ij} measures the strength of connection between contacts i and j."

jimi adams. EPIC- SNA 2017. Columbia University. Origin: Burt RS. Secondhand Brokerage: Evidence on the Importance of Local Structure for Managers, Bankers, and Analysts. Academy of Management Journal 2007;50(1):119-148.

Measuring the Absence of Structural Holes, pt 1

$$C_{ij} = \left(\boxed{p_{ij}} + \boxed{\sum_q p_{iq} p_{qj}} \right)^2$$



C_{ij} = Direct investment (P_{ij}) + Indirect investment

C	1	2	3	4	5
1		.11	.06	.11	.25
2	.44		.02	.08	.39
3	1	.06		.06	.06
4	.44	.08	.02		.39
5	.44	.17	.01	.17	

	C
1	.53
2	.83
3	1
4	.83
5	.78

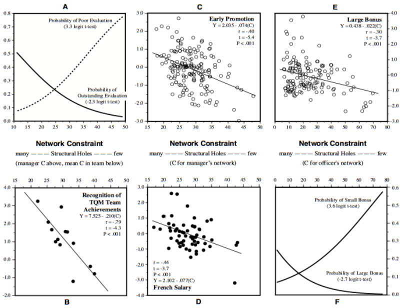
P	1	2	3	4	5
1		.25	.25	.25	.25
2	.5		0	0	.5
3	1	0		0	0
4	.5	0	0		.5
5	.33	.33	0	.33	

P*P	1	2	3	4	5
1		.08	0	.08	.25
2	.17		.13	.29	.13
3	0	.25		.25	.25
4	.17	.29	.13		.13
5	.33	.08	.08	.08	

$P+P^2$	1	2	3	4	5
1		.33	.25	.33	.5
2	.67		.13	.29	.63
3	1	.25		.25	.25
4	.67	.29	.13		.63
5	.67	.41	.08	.41	

Structural Holes

Figure 2. Social Capital Matters



Burt, RS.

Structural Holes versus Network Closure as Social Capital.

Pp. 31-56 in Lin N, Cook K, Burt RS (eds). *Social Capital: Theory and Research.*

Aldine de Gruyter; 2001.

Structural Holes

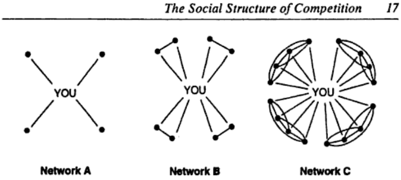
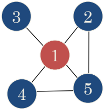


Figure 1.1 Network expansion

jimi adams. EPIC- SNA 2017. Columbia University. Origin: Burt RS. Structural Holes: The social structu

Structural Holes

Measuring the Presence of
Structural Holes, pt 2



E = Network Size - Redundancies

Efficiency = $\frac{\text{Effective Size}}{\text{Size}}$

$$\sum_j \left[1 - \sum_q p_{iq} m_{jq} \right]$$

Node 1

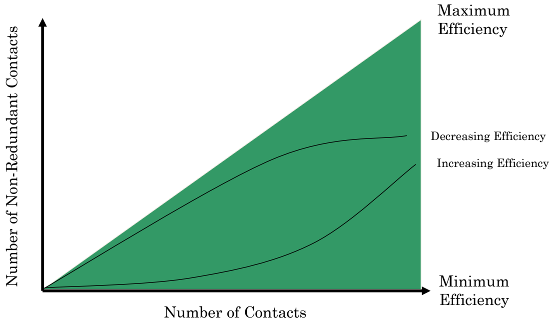
4 – (Σ = 1) = 3

	pm
1	
2	1/4
3	0
4	1/4
5	2/4

Node	**Size**	**Effective Size**	**Efficiency**
1	4	3	.75
2	2	1	.5
3	1	1	1
4	2	1	.5
5	3	1.67	.56

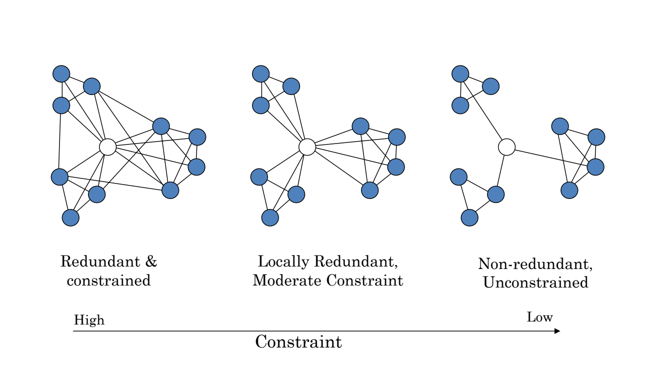
jimi adams. EPIC- SNA 2017. Columbia University. Origin: Burt RS. Structural Holes: The social structure of competition. Harvard University Press; 199

Structural Holes



jimi adams. EPIC- SNA 2017. Columbia University. Origin: Granovetter M. The Strength of Weak Ties. American Journal of Sociology 1973;81:1287-1303.

Structural Holes



jimi adams. EPIC- SNA 2017. Columbia University. Origin: Jim Moody

Structural Holes

TABLE 3—Regression Models Using the Diffusing Agency's Network Position and Structural Characteristics of a Public Health Network to Predict Systemwide Information Diffusion

Variable	Medium Priority[a]		High Priority[b]	
	Model 1[c,d]	Model 2[c,e]	Model 1[c,d]	Model 2[c,e]
Mean partnering tendency	0.364*	0.104*	0.409*	0.087*
Standard deviation in partnering	-0.036	-0.022	-0.056	-0.032*
Fully connected structure		-0.105*		-0.112
Chain structure		-0.810*		-0.870*
Hierarchy structure		-0.396*		-0.479*
Connected clusters structure		-0.252*		-0.282*
Diffuser's effective network		0.429*		0.518*
R^2 model	0.119	0.726	0.144	0.877
F statistic for R^2 (*df*)	202.018* (2, 2997)	1133.179* (7, 2992)	251.081* (2, 2997)	3046.901* (7, 2992)
R^2 change		0.607		0.733
F statistic for change in R^2 (*df* change)		1326.895* (5, 2992)		3567.624* (5, 2992)

Gibbons DE.

Interorganizational Network Structures and Diffusion of Information through a Health System.

American Journal of Public Health 2007;97(9):1684-1692.

Jonas AB, Young AM, Oser CB, Leukefeld CG, Havens JR.

OxyContin® as Currency: OxyContin® use and Increased Social Capital among Appalachian Drug Users

Social Science and Medicine 2012;74:1602-1609.

Variable	AOR	95% CI	*p*-value
Male gender	0.79	0.56–1.10	0.159
Years of education	1.00	0.99–1.01	0.260
Total monthly income	1.00	1.00–1.00	0.212
Daily marijuana use	0.62	0.44–0.87	0.005
Daily alcohol use to intoxication	0.57	0.26–1.25	0.158
Daily hydrocodone use	0.80	0.56–1.14	0.222
Daily OxyContin® use	2.31	1.61–3.30	<0.0001

AOR: adjusted odds ratio, CI: confidence interval.

Distance Measures

Distance Measures (reachability)

(Image Source: Wikipedia)

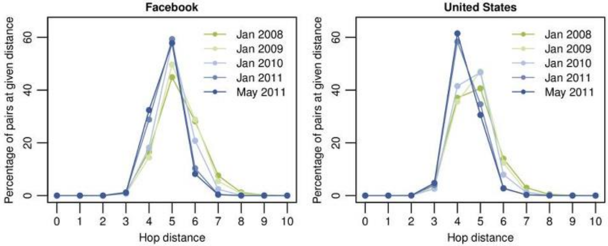
Stanley Milgram (1960s)

- Questions:
  - What are the chances that any two people chosen at random are connected by a path of some length?
  - How long are those chains of intermediaries?
    - Maximally
    - On Average
- Procedure
  - Select "Random" Seeds & Target
  - Ask seeds to mail package to individuals they mutually "know by face & name" closer to target
  - Told city/occupation of target
  - In addition to sending the package, notify research team of each step in the chain
  - Count steps in chains & plot

Distance Measures (reachability)

The Small World of Facebook

- Mean 2008 = 5.28
- Mean 2011 = 4.74



jimi adams. EPIC- SNA 2017. Columbia University.
[Facebook](https://www.facebook.com/notes/facebook-data-team/anatomy-of-facebook/10150388519243859)

Distance Measures (reachability)

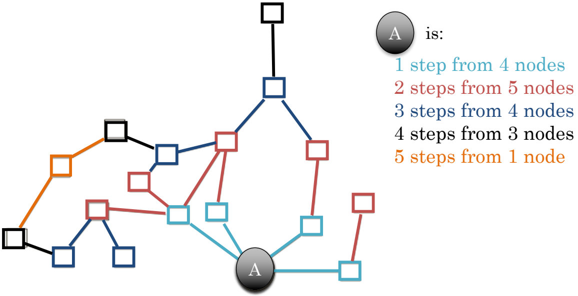
- adjacent – two nodes connected by a line
  aka *directly* connected
- neighborhood – complete set of adjacent nodes
- degree – numerical measure of neighborhood (sans ego)

- walk – sequence of alternating nodes/lines
- trail – a walk consisting of unique arcs/edges
- path – a trail consisting of unique nodes
- circuit – a trail beginning and ending with the same node
- cycle – a path beginning and ending with the same node

- walk/trail/path length – number of *lines* in a w/t/p
- distance (geodesic) - length of shortest path

Geodesics

Geodesic distance is the length of the *shortest* path connecting a pair of nodes.



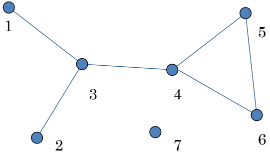
Distance Measures (reachability)

Geodesics

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

GD_M

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

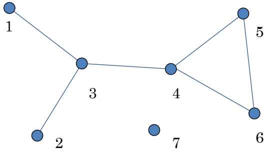


Distance Measures (reachability)

Geodesics

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

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



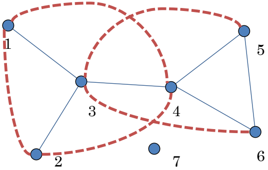
Distance Measures (reachability)

Geodesics

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

M^2	1	2	3	4	5	6	7
1	1	1	0	1	0	0	0
2	1	1	0	1	0	0	0
3	0	0	3	0	1	1	0
4	1	1	0	3	1	1	0
5	0	0	1	1	2	1	0
6	0	0	1	1	1	2	0
7	0	0	0	0	0	0	0

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



jimi adams. EPIC- SNA 2017. Columbia University.

Distance Measures (reachability)

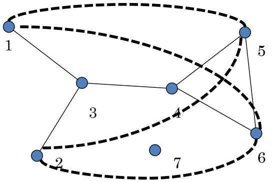
Geodesics

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

M^2	1	2	3	4	5	6	7
1	1	1	0	1	0	0	0
2	1	1	0	1	0	0	0
3	0	0	3	0	1	1	0
4	1	1	0	3	1	1	0
5	0	0	1	1	2	1	0
6	0	0	1	1	1	2	0
7	0	0	0	0	0	0	0

M^3	1	2	3	4	5	6	7
1	0	0	3	0	1	1	0
2	0	0	3	0	1	1	0
3	3	3	0	5	1	1	0
4	0	0	5	2	4	4	0
5	1	1	1	4	2	3	0
6	1	1	1	4	3	2	0
7	0	0	0	0	0	0	0

GD_M	1	2	3	4	5	6	7
1	0	2	1	2	3	3	∞
2	2	0	1	2	3	3	∞
3	1	1	0	1	2	2	∞
4	2	2	1	0	1	1	∞
5	3	3	2	1	0	1	∞
6	3	3	2	1	1	0	∞
7	∞	∞	∞	∞	∞	∞	0



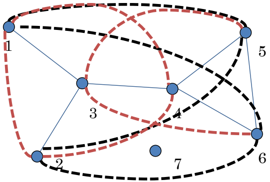
Geodesics

R_M

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

GD_M

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

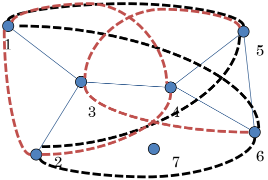


Geodesics

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

GD_M	1	2	3	4	5	6	7
1	0	2	1	2	3	3	∞
2	2	0	1	2	3	3	∞
3	1	1	0	1	2	2	∞
4	2	2	1	0	1	1	∞
5	3	3	2	1	0	1	∞
6	3	3	2	1	1	0	∞
7	∞	∞	∞	∞	∞	∞	0

R_M	1	2	3	4	5	6	7
1	.	1	1	1	1	1	0
2	1	.	1	1	1	1	0
3	1	1	.	1	1	1	0
4	1	1	1	.	1	1	0
5	1	1	1	1	.	1	0
6	1	1	1	1	1	.	0
7	0	0	0	0	0	0	.

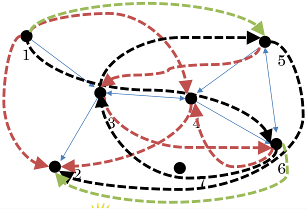


Geodesics

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

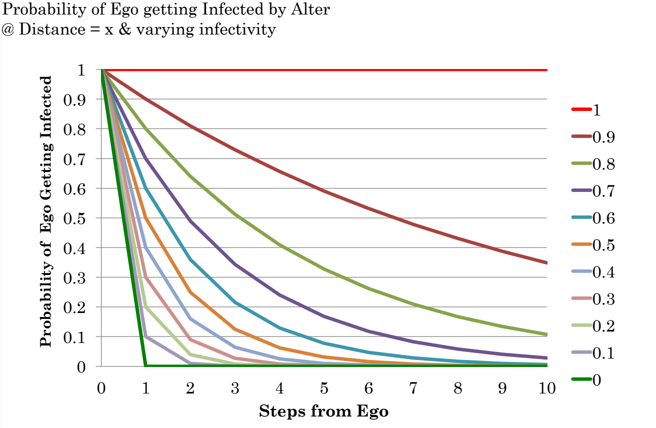
GD M	1	2	3	4	5	6	7
1	0	2	1	2	4	3	∞
2	∞	0	∞	∞	∞	∞	∞
3	∞	1	0	1	3	2	∞
4	∞	2	1	0	2	1	∞
5	∞	3	2	1	0	1	∞
6	∞	4	3	2	1	0	∞
7	∞	∞	∞	∞	∞	∞	0

R_M	1	2	3	4	5	6	7
1	.	1	1	1	1	1	0
2	0	.	0	0	0	0	0
3	0	1	.	1	1	1	0
4	0	1	1	.	1	1	0
5	0	1	1	1	.	1	0
6	0	1	1	1	1	.	0
7	0	0	0	0	0	0	.



Distance Measures (reachability)

Geodesics



jimi adams. EPIC- SNA 2017. Columbia University.

Distance Matters

Example: Obesity Spread

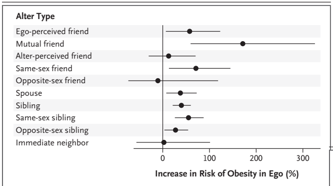
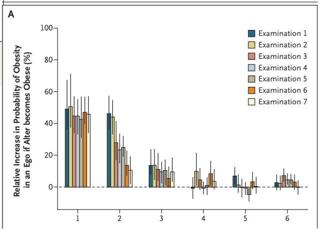


Figure 4. Probability That an Ego Will Become Obese According to the Type of Relationship with an Alter Who May Become Obese in Several Subgroups of the Social Network of the Framingham Heart Study.



Christakis NA, Fowler JH.

The Spread of Obesity in a Large Social Network over 32 Years.

New England Journal of Medicine 2007;357(4):370-379.

Distance Measures (reachability)

Distance Matters

Some additional notes

- Diameter – length of longest shortest path in the network
- What to do with unconnected pairs?
  - Technically if reach = 0, distance = ∞
    - Could be measurement error
    - There's a reason you included them in the network, so infinity doesn't make theoretical sense
    - Moreover, mathematical manipulations w/∞ are frequently intractable
  - What to use?
    - Network-size based:
      - N-1 maximum theoretical distance **(most common)**
      - N max distance +1 (avoids distance for disconnected pairs = connected)
    - Diameter + 1
      - Unreachable pairs one step farther apart than most distant ones

Incomplete Path Lengths

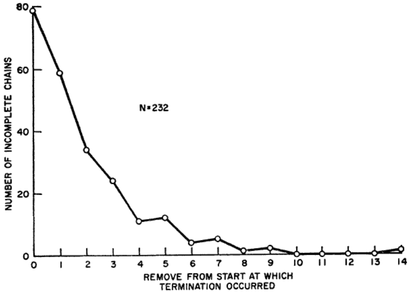


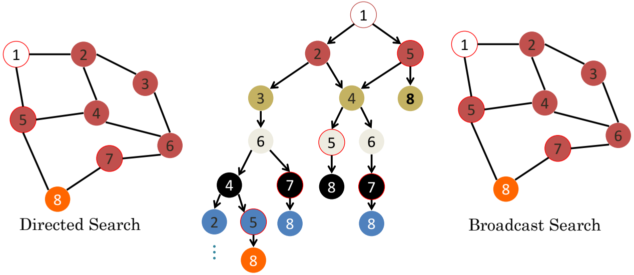
FIGURE 2

Lengths of Incomplete Chains

jimi adams. EPIC- SNA 2017. Columbia University.

Is 6 large or small number?

- Actual vs. "Geodesic" path
- There's a mismatch between the theory and the test



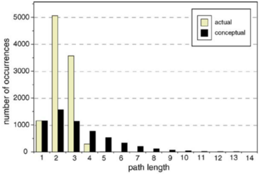
jimi adams. EPIC- SNA 2017. Columbia University.

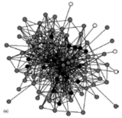
Distance Measures (reachability)

Direct Searches

How Badly do They Fail?

- Map social networks within a telecom firm
  - Compute *actual* geodesic distances between all pairs of nodes
- Ask the employees to enact an internal “small world” experiment
  - Compare efficiency of “conceptual” paths to actual geodesics



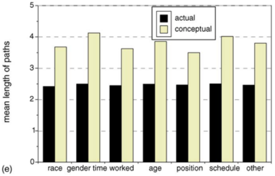


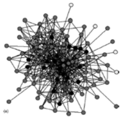
- How well did the “conceptual” paths do?
  - ~50% longer
  - Only 32% were “accurate”
  - Chose wrong next step ½ the time

Distance Measures (reachability)

Why were they so far off?

- People use simplifying assumptions
  - We group people into identifiable traits
  - Assume passing to someone "more like" the target on those characteristics will be closer to them
  - We typically only use one (or a small number) of characteristics to "simplify" on at a time





jimi adams. EPIC- SNA 2017. Columbia University. Origin: Killworth PD, McCarty C, Bernard HR, House M. The Accuracy of Small World Chains in Social Networks. Social Networks 2006;28.

Distance Measures (reachability)

Female	⟶	Female	56
Male	⟶	Male	58
Female	⟶	Male	18
Male	⟶	Female	13

56	18	74
13	58	71
69	76	

35	39	74
34	37	71
69	76	

$E(X) = R*C/T$

$$Seg = \frac{E(X) - X}{E(X)}$$

$E(X) = (39+34)$
$X = (18+13)$
$Seg = (73-31)/\ 73 = 0.575$**

jimi adams. EPIC- SNA 2017. Columbia University. Origin: Milgram S. The Small World Problem. Psychology Today 1967;1:61-6.}

Distance Measures (reachability)

Milgram's Experiments

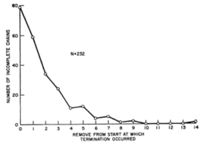
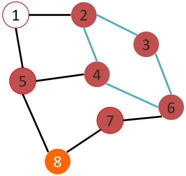


FIGURE 2

Lengths of Incomplete Chains



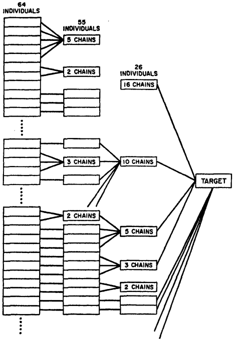


FIGURE 3

Common Paths Appear as Chains Converge on the Target

jimi adams. EPIC- SNA 2017. Columbia University. Origin: Milgram S. The Small World Problem. Psychology Today 1967;1:61-6

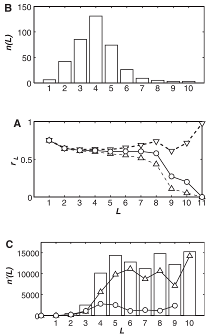
Distance Measures (reachability)

Table 1. Type, origin, and strength of social ties used to direct messages. Only the top five categories in the first two columns have been listed. The most useful category of social tie is medium-strength friendships that originate in the workplace.

Type of relationship	%	Origin of relationship	%	Strength of relationship	%
Friend	67	Work	25	Extremely close	18
Relatives	10	School/university	22	Very close	23
Co-worker	9	Family/relation	19	Fairly close	33
Sibling	5	Mutual friend	9	Casual	22
Significant other	3	Internet	6	Not close	4

Table 2. Reason for choosing next recipient. All quantities are percentages. Location, recipient is geographically closer; Travel, recipient has traveled to target's region; Family, recipient's family originates from target's region; Work, recipient has occupation similar to target; Education, recipient has similar educational background to target; Friends, recipient has many friends; Cooperative, recipient is considered likely to continue the chain; Other, includes recipient as the target.

L	*N*	Location	Travel	Family	Work	Education	Friends	Cooperative	Other
1	19,718	33	16	11	16	3	9	9	3
2	7,414	40	11	11	19	4	6	7	2
3	2,834	37	8	10	26	6	6	4	3
4	1,014	33	6	7	31	8	5	5	5
5	349	27	3	6	38	12	6	3	5
6	117	21	3	5	42	15	4	5	5
7	37	16	3	3	46	19	8	5	0



jimi adams. EPIC- SNA 2017. Columbia University. Origin: Milgram S. The Small World Problem. Psychology Today 1967;1:61-6

R for Descriptive Analysis: More tools

Degree Distribution

Network Composition

Network Position (Theories)

Distance Measures

R for Descriptive Analysis: More tools