Social Network Analysis – Network Dynamics

The Small-world Phenomenon

Overview

- Six Degrees of Separation
- Structure and Randomness
- Decentralized Search
- Generalized Models
- Core-Periphery Structures

Overview

- Six Degrees of Separation
- Structure and Randomness
- Decentralized Search
- Generalized Models
- Core-Periphery Structures

Small-world Phenomenon

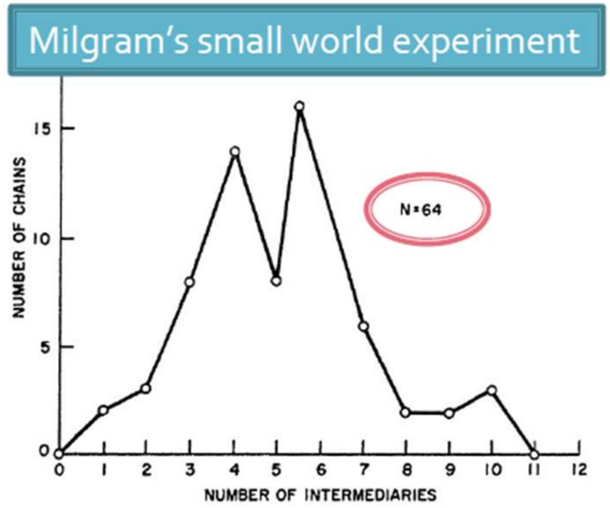
- Most Social Networks have a giant connected component – there are paths of friends connecting you to a large fraction of the world's population
- These paths are surprisingly short.
- This idea has been termed the small-world phenomenon — the idea that the world looks "small" when you think of how short a path of friends it takes to get from you to almost anyone else.
- The fact that social networks are so rich in short paths is known as the *small-world phenomenon*, or the "*six degrees of separation*"
  - Has long been the subject of both anecdotal and scientific fascination.

The Small-world Experiment

- The first significant empirical study of the small-world phenomenon was undertaken by the social psychologist Stanley Milgram in the 1960s [2]
  - Picked 296 people at random
  - Ask them to get a letter to a by passing it through friends to a stockbroker in Boston
  - He provided the target's name, address, occupation, and some personal information, but stipulated that the participants could not mail the letter directly to the target;
  - Rather, each participant could only advance the letter by forwarding it to a single acquaintance that he or she knew on a first-name basis, with the goal of reaching the target as rapidly as possible.

The Small-world Experiment

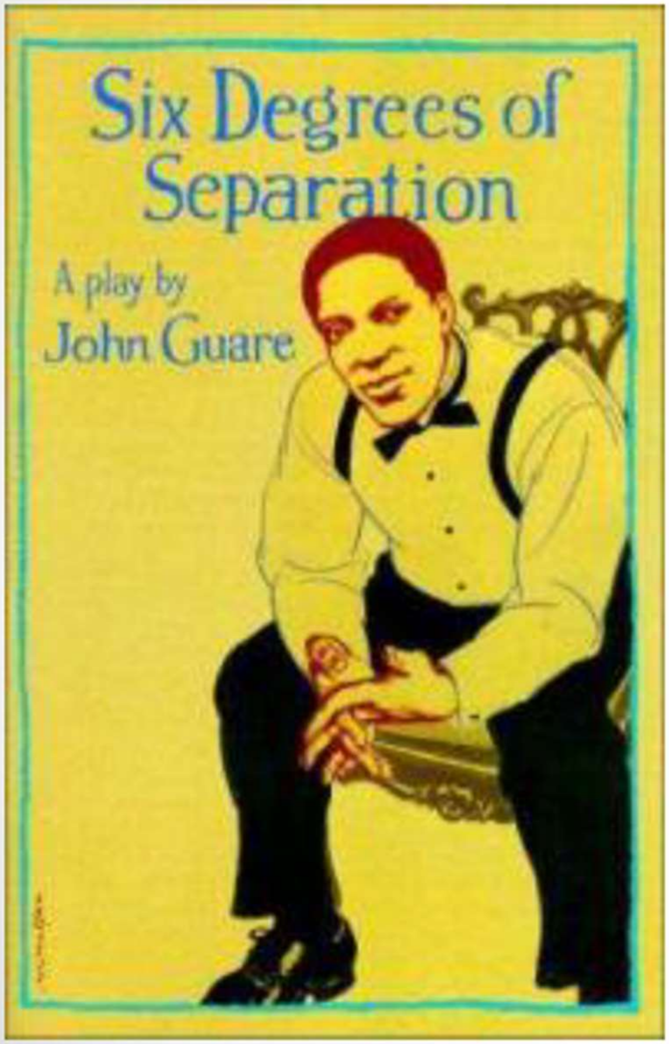
- 64 chains completed:
  - 6.2 on the average, thus "6 degrees of separation"
- Further observations:
  - People what owned stock had shortest paths to the stockbroker than random people
  - People from the Boston area have even closer paths



Milgram's experiment really demonstrated two striking facts about large social networks:

1. Short paths are there in abundance
2. People, acting without any sort of global "map" of the network, are effective at collectively finding these short paths

Six degrees of Separation

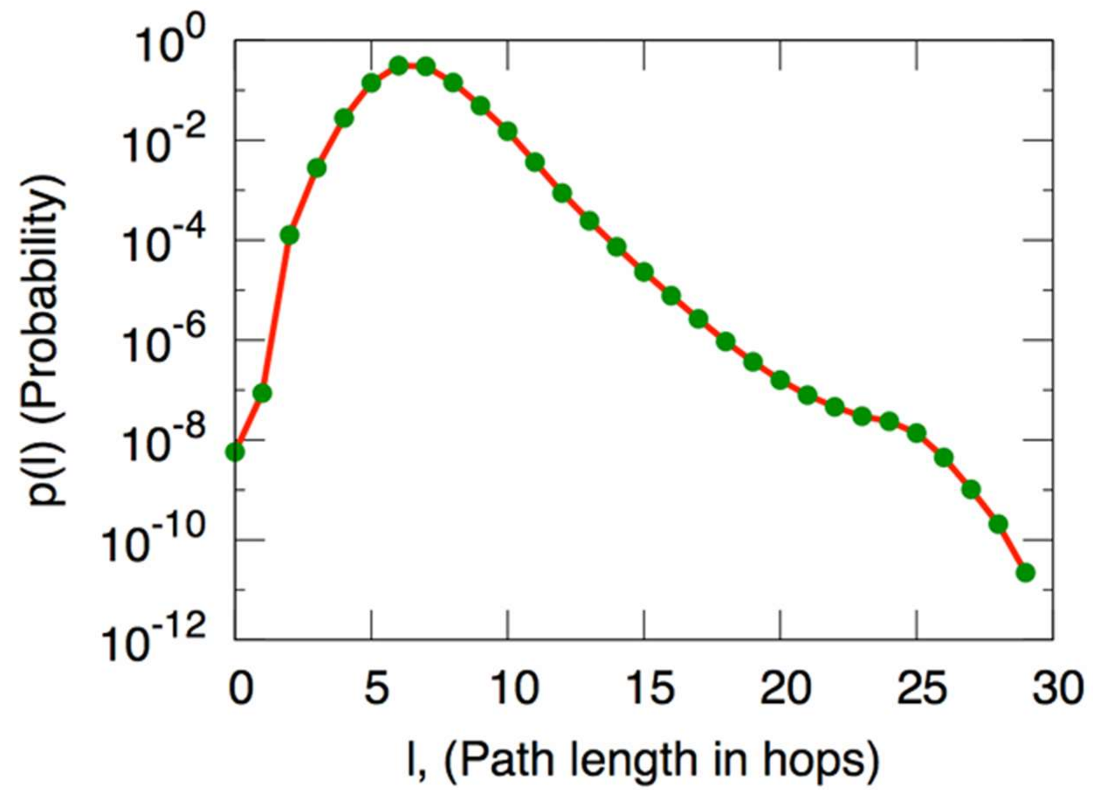


"Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice…. It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought. How every person is a new door, opening up into other worlds." [1]

Instant Messaging Experiment

- Jure Leskovec and Eric Horvitz [3] analyzed the 240 million active user accounts on Microsoft Instant Messenger
- Built a graph:
  - Node corresponds to a user
  - There is an edge between two users if they engaged in a two-way conversation at any point during a month-long observation period.
- This graph turned out to have a giant component containing almost all of the nodes
  - The distances within this giant component were very small.
  - The distances in the Instant Messenger network closely corresponded to the numbers from Milgram's experiment, with an estimated average distance of 6.6, and an estimated median of seven.

Instant Messaging Experiment

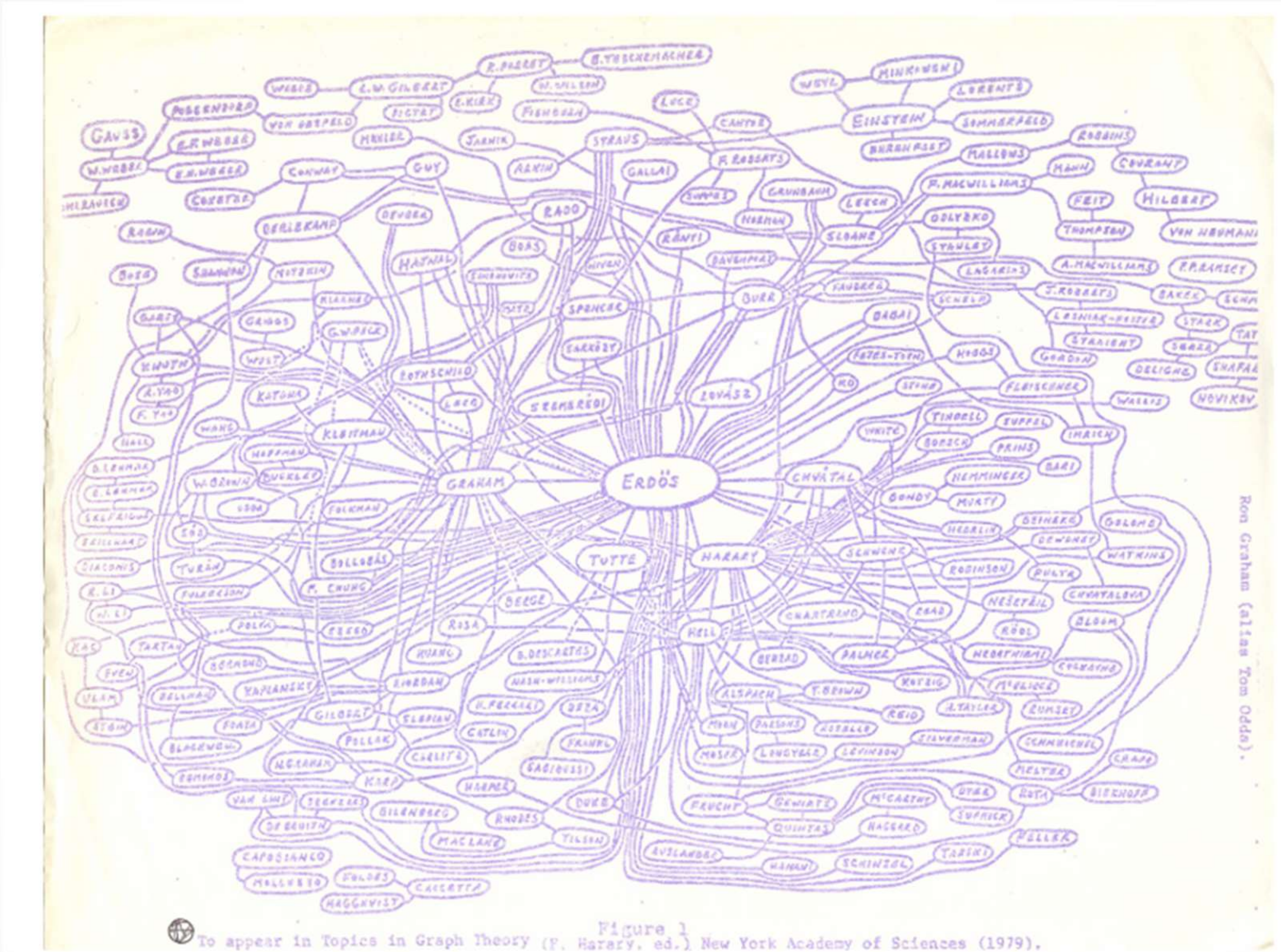


The distribution of distances in the graph of all active Microsoft Instant Messenger user accounts averaged over a random sample of 1000 users

Small-world in Collaboration Networks

- Researchers have also discovered very short paths in the collaboration networks within professional communities.
- In the domain of mathematics, for example, people often speak of the itinerant mathematician Paul Erdos — who published roughly 1500 papers over his career — as a central figure in the collaborative structure of the field.
- To make this precise, we can define a collaboration graph with nodes corresponding to mathematicians, and edges connecting pairs who have jointly authored a paper.
- Now, a mathematician's Erdos number is the distance from him or her to Erdos in this graph [4].
- Most mathematicians have Erdos numbers of at most 4 or 5
- Extending the collaboration graph to include co-authorship across all the sciences — most scientists in other fields have Erdos numbers that are comparable or only slightly larger
    - Albert Einstein's is 2
    - Enrico Fermi's is 3
    - Noam Chomsky ia 4
    - Francis Crick's and James Watson's are 5 and 6 respectively.

Erdos Number

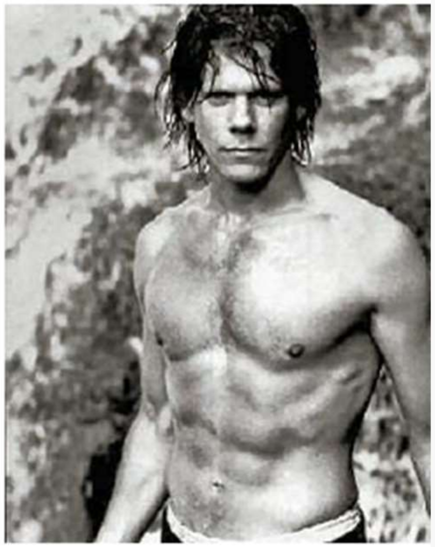


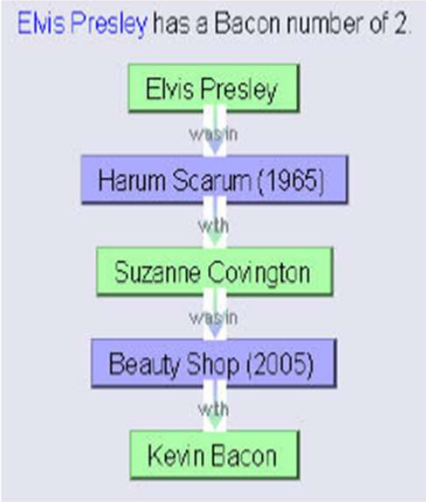
Ron Graham's hand-drawn picture of a part of the mathematics collaboration graph, centered on Paul Erdos [5]

Six degrees of Kevin Bacon

Three students at Albright College in Pennsylvania sometime around 1994 adapted the idea of Erdos numbers to the collaboration graph of movie actors and actresses using IMDB [6]

- Bacon number:
  - Create a network of Hollywood actors
  - Connect two actors if they co-appeared in the movie
  - Bacon number: number of steps to Kevin Bacon
- As of Dec 2007, the highest (finite) Bacon number reported is 8
- Only approx. 12% of all actors cannot be linked to Bacon

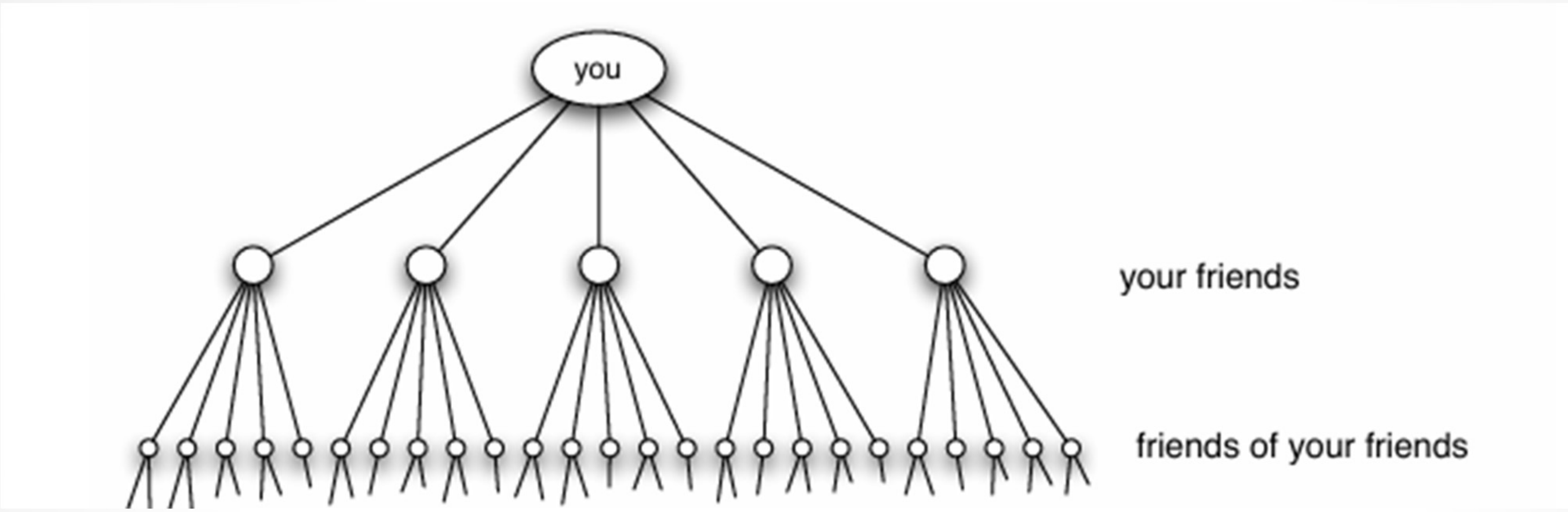




Overview

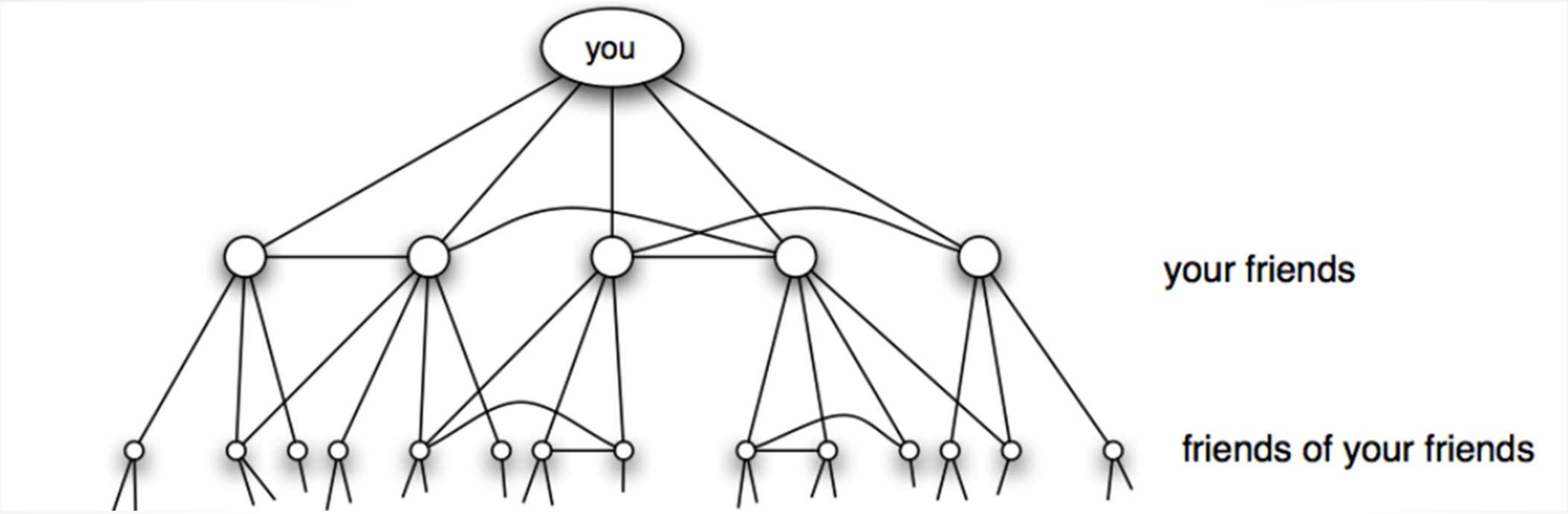
- Six Degrees of Separation
- Structure and Randomness
- Decentralized Search
- Generalized Models
- Core-Periphery Structures

Existence of Short Paths



If each of us know 100 people, the numbers are growing by powers of 100 with each step, bringing us to 100 million after four steps, and 10 billion after five steps
Pure exponential growth produces a small world and short paths

Existence of Short Paths



The effect of triadic closure in social networks works to limit the number of people you can reach by following short paths

2 conflicting features

- Perhaps, the small-world phenomenon surprising to many people because:
  - from the local perspective of any one individual the social network appears to be highly clustered
  - not the kind of massively branching structure that would lead to short paths

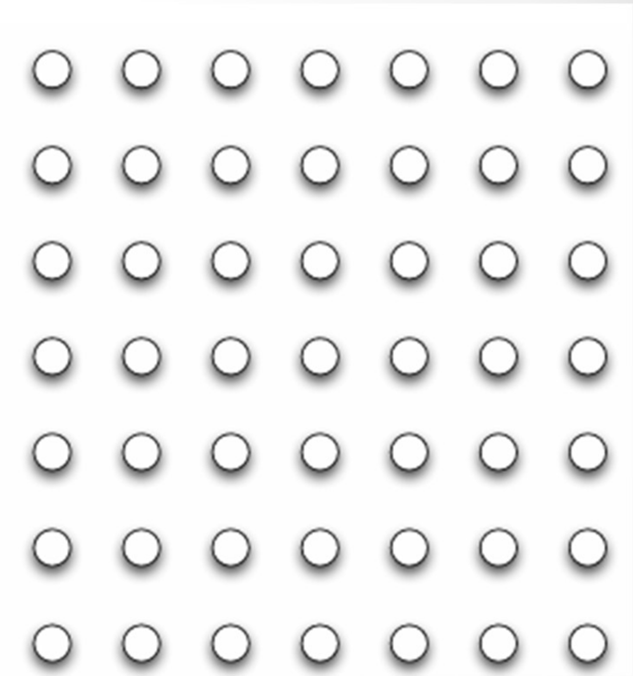
- Can we make up a simple model that exhibits both of the features we've been discussing:
  - many closed triads,
  - but also very short paths?

Watts-Strogatz Model

- Watts and Strogatz [7] proposed a very simple model that generates random networks with the desired properties
- Follows naturally from combination of two basic social-network concepts:
  - **Homophily** (the principle that we connect to others who are like ourselves): creates many triangles
  - **Weak ties** (the links to acquaintances that connect us to parts of the network that would otherwise be faraway): produce the kind of widely branching structure that reaches many nodes in a few steps producing short paths

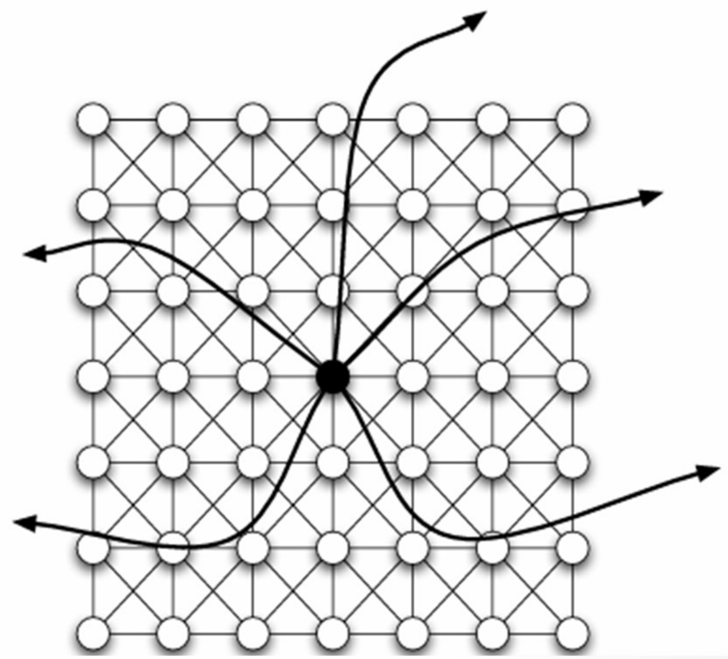
Watts-Strogatz Model

- Let's suppose that everyone lives on a two-dimensional grid
- The grid is a model of geographic proximity, or potentially some more abstract kind of social proximity.
- It is a notion of similarity that guides the formation of links.
- Two *nodes are one grid step apart if they are directly* adjacent to each other in either the horizontal or vertical direction
- Homophily is captured by having each node form a link to all other nodes that lie within a radius of up to *r* grid steps away, for some constant value of r
- These are the links you form to people because you are similar to them.



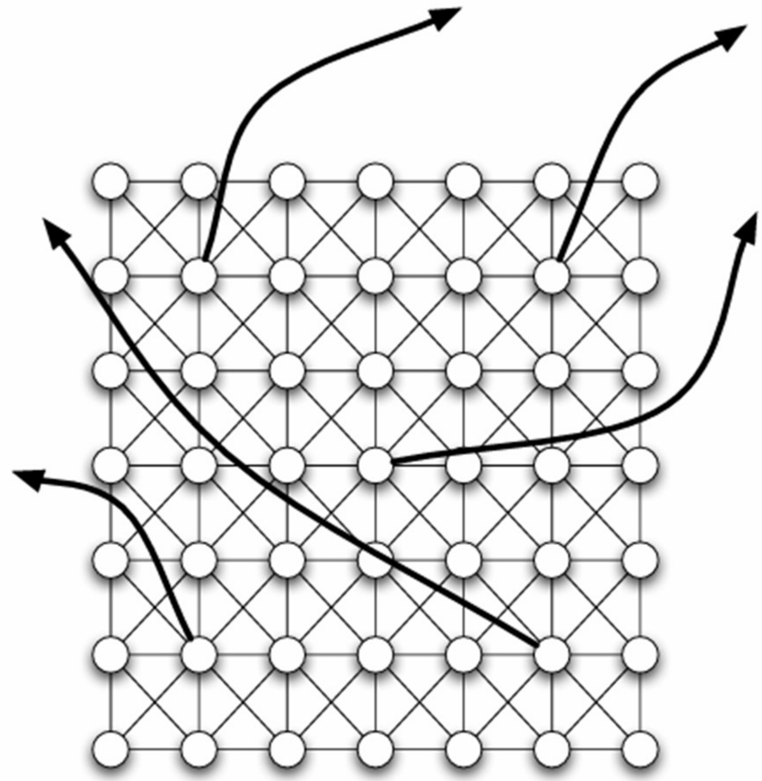
Watts-Strogatz Model

- For some other constant value *k*, each node also forms a link to k other nodes selected uniformly at random from the grid.
- These correspond to weak ties, connecting nodes who lie very far apart on the grid.
- This is a hybrid structure consisting of a small amount of randomness (the weak ties) sprinkled onto an underlying structured pattern (the homophilous links).
- Watts and Strogatz observed first that the network has many triangles: any two neighboring nodes (or nearby nodes) will have many common friends, where their neighborhoods of radius *r* overlap, and this produces many triangles.
- But they also find that there are — with high probability — very short paths connecting every pair of nodes in the network.



Watts-Strogatz Model

- Suppose we only allow one out of every *k* nodes to have a single random friend — keeping the proximity based edges as befor
- Even this network will have short paths between all pairs of nodes.
- Crux of the Watts-Strogatz model: introducing a tiny amount of randomness - in the form of long-range weak ties — is enough to make the world "small," with short paths between every pair of nodes



Overview

- Six Degrees of Separation
- Structure and Randomness
- Decentralized Search
- Generalized Models
- Core-Periphery Structures

Milgram's Experiment: Decentralized Search

- The second basic aspect of the Milgram small-world experiment — the fact that people were actually able to collectively find short paths to the designated target.
- This novel kind of "social search" task was a necessary consequence of the way Milgram formulated the experiment for his participants - with the letter advancing just one person at a time
- The success of the experiment raises fundamental questions about the power of collective search:
  - Even if the social network contains short paths, what made this type of decentralized search so effective?

Milgram's Experiment: Decentralized Search

- Can we use the Walter-Strogatz model to explain this?
  - Initially starting node s only knows the location of destination node t on the grid, but, crucially, it does not know the random edges out of any node other than itself.
  - Each intermediate node along the path has this partial information as well, and it must choose which of its neighbors to send the message to next.
  - These choices amount to a collective procedure for finding a path from s to t — just as the participants in the Milgram experiment collectively constructed paths to the target person.

Watts-Strogatz Model - Limitation

- Unfortunately, one can prove that decentralized search in the Watts Strogatz model will necessarily require a large number of steps to reach a target—much larger than the true length of the shortest path [8].
- The Watts Strogatz network is effective at capturing the density of triangles and the existence of short paths
- But not the ability of people, working together in the network, to actually find the paths.
- Essentially, the problem is that the weak ties that make the world small world are "too random" in this model:
- They're completely unrelated to the similarity among nodes that produces the homophily-based links, they're hard for people to use reliably.

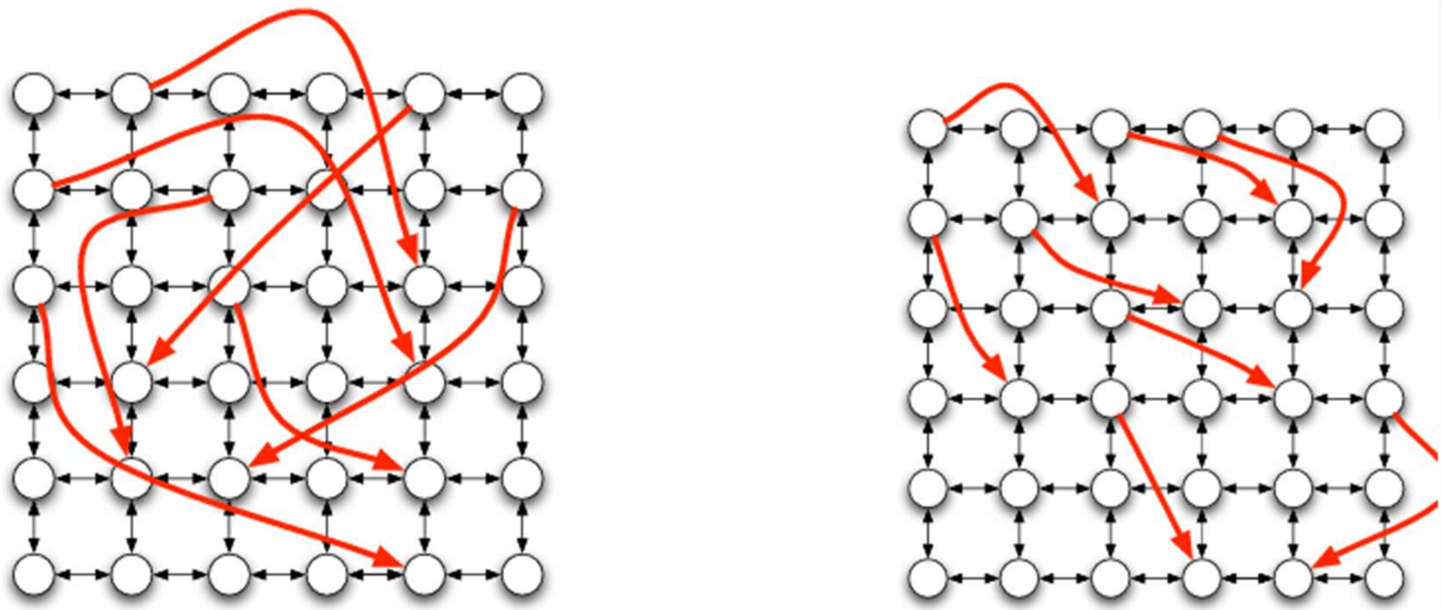
Generalizing the Watts-Strogatz Model

- A mild generalization of the Watts Strogatz model developed by Kleinberg[8] exhibits both properties:
  i. the networks contain short paths
  ii. these short paths can be found using decentralized search
- We introduce one extra quantity that controls the "scales" spanned by the long-range weak ties.
- We have nodes on a grid as before, and each node still has edges to each other node within r grid steps.
- But now, each of its *k* random edges is generated in a way that decays with distance, controlled by a clustering exponent *q* as follows.
  - For two nodes v and w, let d(v,w) denote the number of grid steps between them.
  - This is their distance if one had to walk along adjacent nodes on the grid.
  - In generating a random edge out of v, we have this edge link to w with probability proportional to $d(v,w)^{-q}$

Generalizing the Watts-Strogatz Model

- Varying *q* is like turning a knob that controls how uniform the random links are.
- We have a different model for each value of *q:*
  I. The original grid-based model corresponds to q = 0, since then the links are chosen uniformly at random
  II. When *q* is very small, the long-range links are "too random," and can't be used e effectively for decentralized search
  III. When *q* is large, the long-range links are "not random enough," since they simply don't provide enough of the long-distance jumps that are needed to create a small world.
- Is there an optimal operating point for the network, where the distribution of long-range links is sufficiently balanced between these extremes to allow for rapid decentralized search?

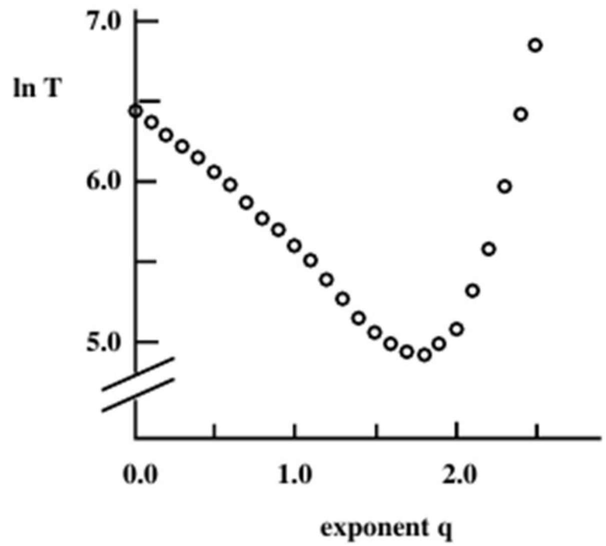
Variations in Clustering Exponent



(a) *A small clustering exponent* (b) *A large clustering exponent*

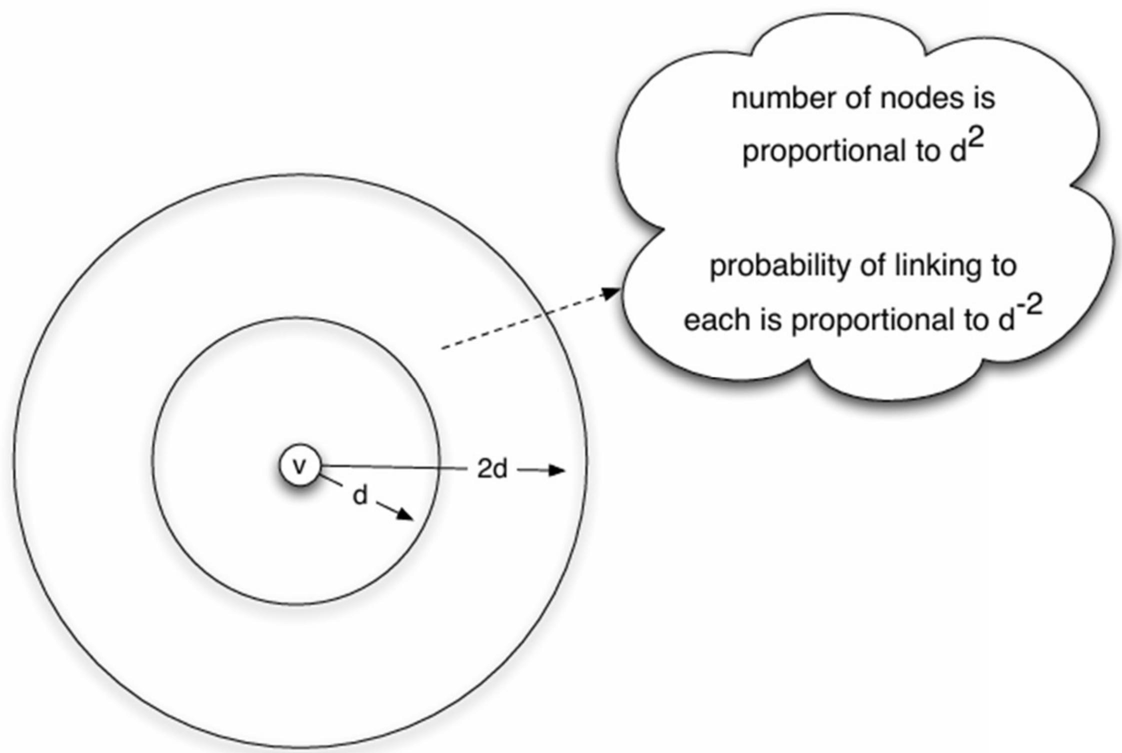
- With a small clustering exponent, the random edges tend to span long distances on the grid
- As the clustering exponent increases, the random edges become shorter.

Optimal Value of Clustering Exponent



- Simulation of decentralized search in the grid-based model with different values of clustering exponent *q*.
- Each point is the average of 1000 runs on (a slight variant of) a grid with 400 million nodes.
- The delivery time is best in the vicinity of exponent q = 2

Why q=2?



- The concentric scales of resolution around a particular node.
- At q=2 the two terms — the number of nodes in the group, and the probability of linking to any one of them — approximately cancel out
- Thus the probability that a random edge links into some node in this ring is approximately independent of the value of d.
- Therefore long-range weak ties are being formed in a way that's spread roughly uniformly over all different scales of resolution.
- This allows people fowarding the message to consistently find ways of reducing their distance to the target, no matter how near or far they are from it

Overview

- Six Degrees of Separation
- Structure and Randomness
- Decentralized Search
- Generalized Models
- Core-Periphery Structures

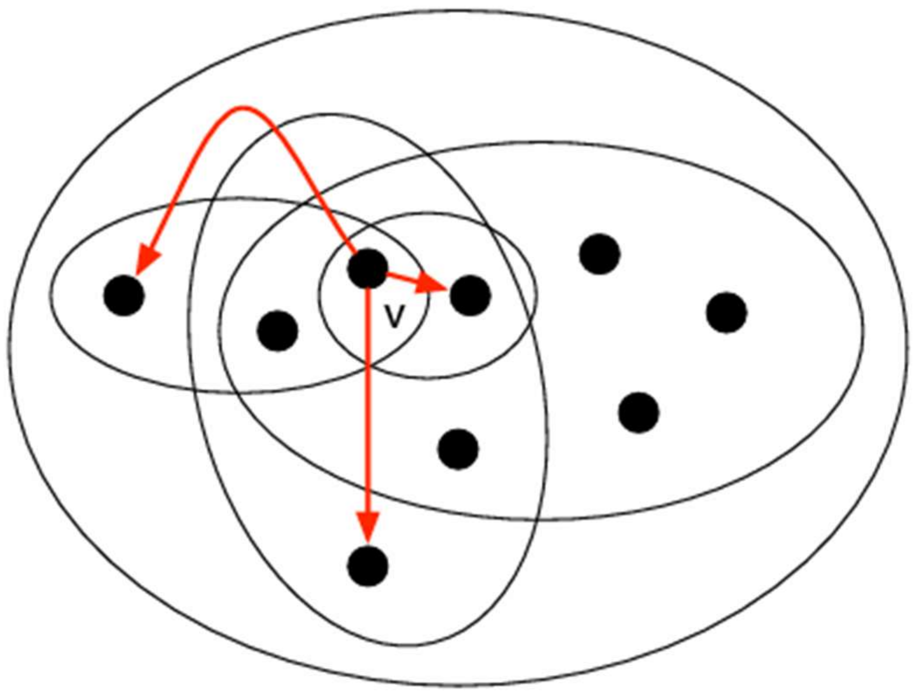
Social Foci & Social Distance

- The notion of social foci provides a flexible and general way to produce models of networks exhibiting both an abundance of short paths and efficient decentralized search.
- A social focus is any type of community, occupational pursuit, neighbor hood, shared interest, or activity that serves to organize social life around it.
- Foci are a way of summarizing the many possible reasons that two people can know each other or become friends: because they live on the same block, work at the same company, frequent the same cafe, or attend the same kinds of concerts.
- Now, two people may have many possible foci in common, but all else being equal, it is likely that the shared foci with only a few members are the strongest generators of new social ties.
- For example, two people may both work for the same thousand-person company and live in the same million-person city, but it is the fact that they both belong to the same twenty-person literacy tutoring organization that makes it most probable they know each other.
- Thus, a natural way to define the social distance between two people is to declare it to be the size of the smallest focus that includes both of them

Social Foci & Social Distance

- Suppose we have a collection of nodes, and a collection of foci they belong to — each focus is simply a set containing some of the nodes.
- We let dist(v, w) denote the social distance between nodes v and w:
  - dist(v, w) is the size of the smallest focus that contains both v and w.
- Let's construct a link between each pair of nodes v and w with probability proportional to $\text{dist}(v,w)^{-p}$.
- One can show, subject to some technical assumptions on the structure of the foci, that when links are generated this way with exponent p = 1, the resulting network supports efficient decentralized search with high probability as shown by Kleinberg[9]

Social Foci & Social Distance

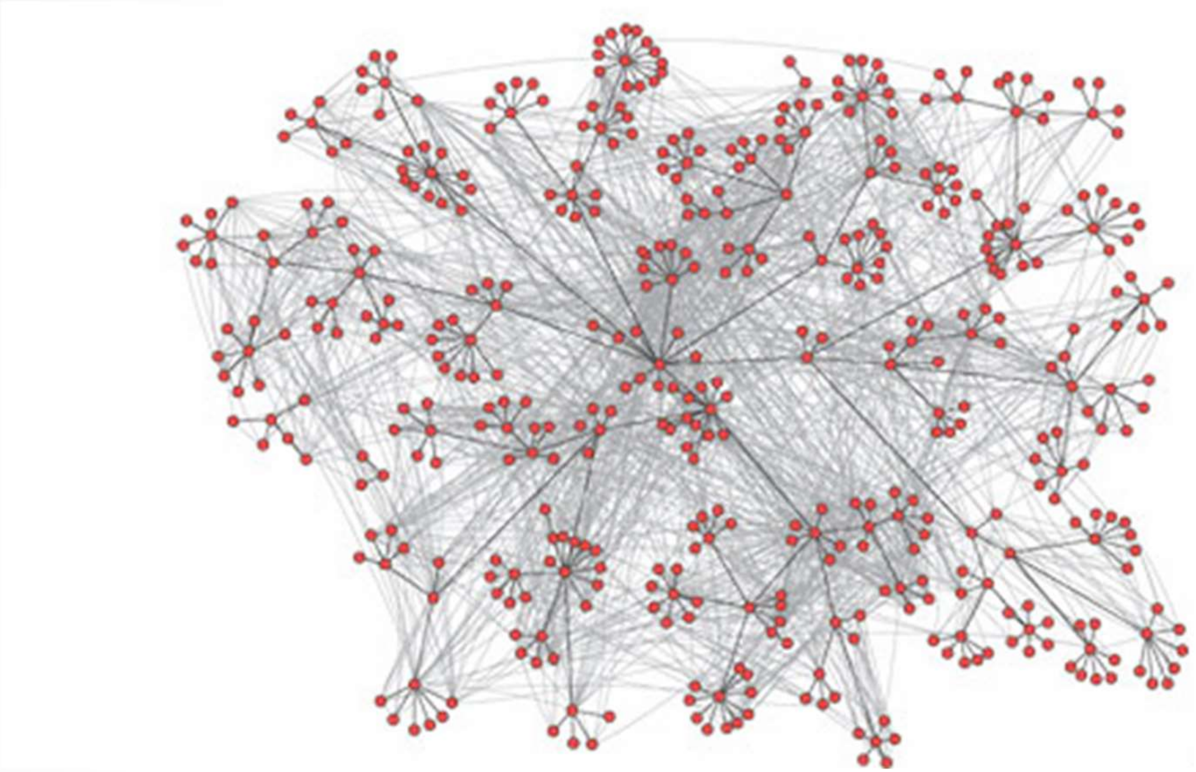


- When nodes belong to multiple foci, we can define the social distance between two nodes to be the smallest focus that contains both of them.
- In the figure, the foci are represented by ovals;
- The node labeled v belongs to five foci of sizes 2,3,5,7, and 9 (with the largest focus containing all the nodes shown).

Social Foci & Social Distance

- Adamic and Adar analyzed a social network connecting two people if they exchanged e-mail at least six times over a three-month period [10].
- They then defined a focus for each of the groups within the organizational structure (i.e. a group of employees all reporting a common manager).
- They found that the probability of a link between two employees at social distance d within the organization scaled proportionally to $d^{-3/4}$.

Social Foci & Social Distance



The pattern of e-mail communication among 436 employees of Hewlett Packard Research Lab is superimposed on the official organizational hierarchy, showing how network links span different social foci

Overview

- Six Degrees of Separation
- Structure and Randomness
- Decentralized Search
- Generalized Models
- Core-Periphery Structures

Recreating Milgram Experiment

- The research community has come to appreciate both the robustness and the delicacy of the “six degrees” principle.
- Many studies of large-scale social network data have confirmed the pervasiveness of very short paths in almost every setting.
- On the other hand, the ability of people to find these paths from within the network is a subtle phenomenon: it is striking that it should happen at all, and the conditions that facilitate it are not fully understood.
- The success rate at finding targets in recreations of the Milgram experiment has often been much lower [11]
- Much of the difficulty can be explained by lack of participation
- But there are also more fundamental difficulties at work, pointing to questions about large social networks that may help inform a richer understanding of network structure.

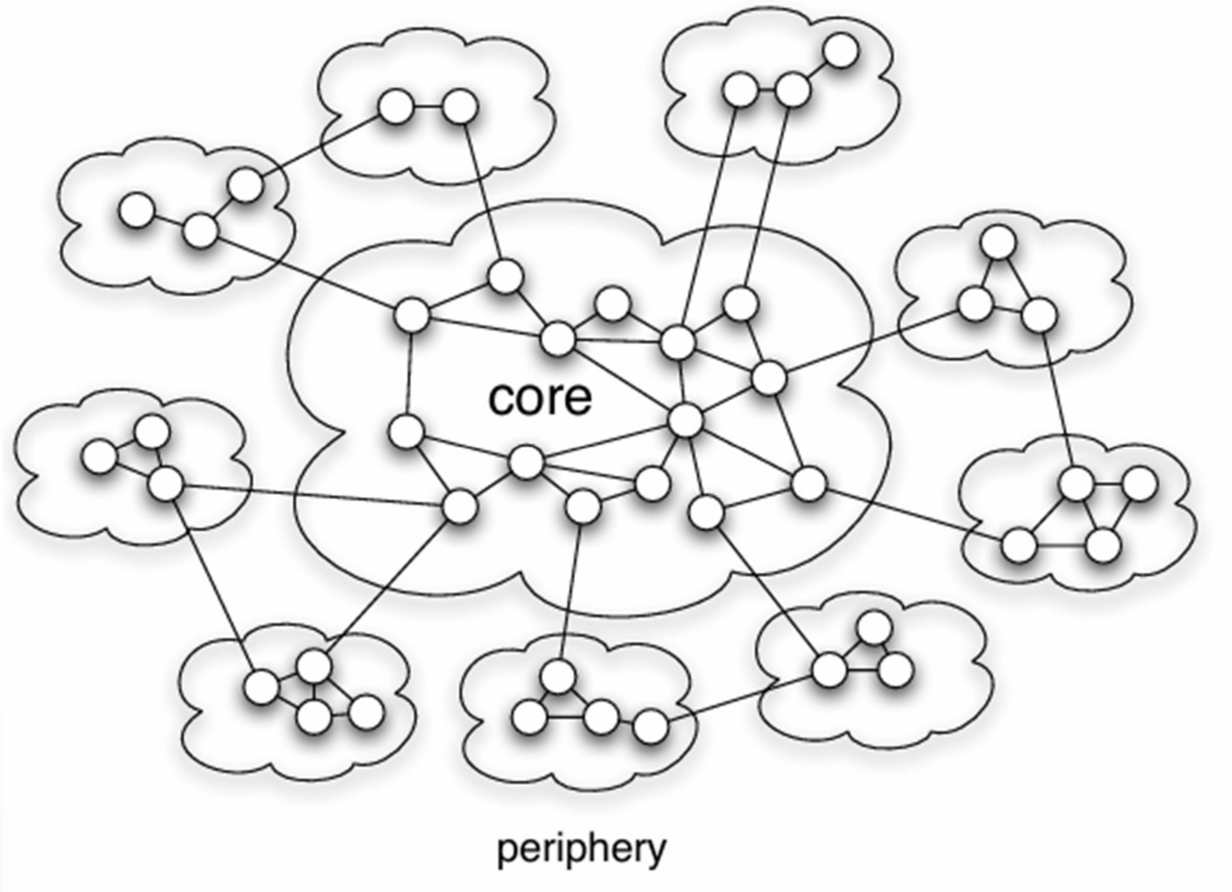
Recreating Milgram Experiment

- Milgram-style search in a network is most successful when the target person is affluent and socially high-status.
- For example, in the largest small-world experiment to date [12], 18 different targets were used, drawn from a wide range of backgrounds.
  - Completion rates to all targets were small, due to lack of participation in the e-mail based forwarding of messages,
  - They were highest for targets who were college professors and journalists
  - Small for low-status targets.
- This wide variation in the success rates of search to different targets does not simply arise from variations in individual attributes of the respective people
  - it is based on the fact that social networks are structured to make high-status individuals much easier to find than low-status ones.

Core-Periphery structure

- Homophily suggests that high-status people will mainly know other high-status people and low-status people will mainly know other low-status people
- This does not imply that the two groups occupy symmetric or interchangeable positions in the social network.
- Large social networks tend to be organized in what is called a core-periphery structure [72], in which the high-status people are linked in a densely-connected core, while the low-status people are atomized around the periphery of the network [13]
- High status people have the resources to travel widely; to meet each other through shared foci around clubs, interests, and educational and occupational pursuits; and more generally to establish links in the network that span geographic and social boundaries.
- Low-status people tend to form links that are much more clustered and local.
- As a result, the shortest paths connecting two low-status people who are geographically or socially far apart will tend to go into the core and then come back out again.

Core-Periphery structure



Core-Periphery structure - Implications

- All this has clear implications for people's ability to find paths to targets in the network.
- In particular, it indicates some of the deep structural reasons why it is harder for Milgram style decentralized search to find low-status targets than high-status targets.
- As you move toward a high-status target, the link structure tends to become richer, based on connections with an increasing array of underlying social reasons.
- In trying to find a low-status target, on the other hand, the link structure becomes structurally more impoverished as you move toward the periphery.

Future Research

- These considerations suggest an opportunity for richer models that take status effects more directly into account.
- The models we have seen capture the process by which people can find each other when they are all embedded in an underlying social structure, and motivated to continue a path toward a specific destination.
- But as the social structure begins to fray around the periphery, an understanding of how we find our way through it has the potential to shed light not just on the networks themselves, but on the way that network structure is intertwined with status and the varied positions that different groups occupy in society as a whole.

References (1/2)

1. John Guare. Six Degrees of Separation: A Play. Vintage Books, 1990.
2. Jerey Travers and Stanley Milgram. An experimental study of the small world problem. Sociometry, 32(4):425{443, 1969.
3. Jure Leskovec and Eric Horvitz. Worldwide buzz: Planetary-scale views on an instant messaging network. In Proc. 17th International World Wide Web Conference, 2008.
4. Jerrold W. Grossman and Patrick D. F. Ion. On a portion of the well-known collaboration graph. Congressus Numerantium, 108:129{131, 1995.
5. Ronald L. Graham. On properties of a well-known graph, or, What is your Ramsey number? Annals of the New York Academy of Sciences, 328(1):166{172, June 1979
6. Simon Singh. Erdos-Bacon numbers. Daily Telegraph, April 2002.
7. Duncan J. Watts and Steven H. Strogatz. Collective dynamics of `small-world' networks. Nature, 393:440-442, 1998.
8. Jon Kleinberg. Navigation in a small world. Nature, 406:845, 2000.

References (2/2)

9. Jon Kleinberg. Small-world phenomena and the dynamics of information. In Proc. 14th Advances in Neural Information Processing Systems, pages 431{438, 2001.
10. Lada A. Adamic and Eytan Adar. How to search a social network. Social Networks, 27(3):187{203, 2005.
11. Judith Kleinfeld. Could it be a big world after all? The `six degrees of separation' myth. Society, 39(2):61{66, January 2002
12. Peter Dodds, Roby Muhamad, and Duncan Watts. An experimental study of search in global social networks. Science, 301:827{829, 2003
13. Stephen P. Borgatti and Martin G. Everett. Models of core/periphery structures. Social Networks, 21(4):375{395, October 2000.

Reading

1. David Easley and Jon Kleinberg. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. https://www.cs.cornell.edu/home/kleinber/networks-book/
   - Chapter 2.3
   - Chapter 20.1-20.5
2. Social Network Analysis. Tanmoy Chakraborty.
   - Chapter 3