

The “New” Science of Networks



Duncan Watts
Yahoo! Research
Columbia University

Outline

- The Small World Problem
 - Aka “Six Degrees of Separation”
- Small World Networks
- “Network Science”
- Some examples of why it matters
 - Disease Spreading and Epidemics
 - Social Influence
 - Networks and Networking
- Challenges and Opportunities



How “Small” is the World?

- “Six degrees of separation between us and everyone else on this planet”
 - John Guare, 1990
- An urban myth?
- First mentioned in 1920’s by Karinthy
- 1950’s Pool and Kochen first posed it as a math problem involving network structure
- First became famous in 1960’s as a result of an ingenious experiment



The Small World Experiment

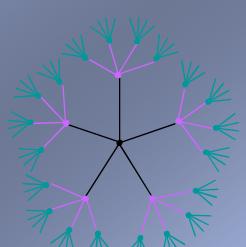
- Stanley Milgram (and student Jeffrey Travers) designed an experiment based on Pool and Kochen’s work
 - A single “target” in Boston
 - 300 initial “senders” in Boston and Omaha
 - Each sender asked to forward a packet to a friend who was “closer” to the target
 - The friends got the same instructions
- Protocol generated 300 “letter chains” of which 64 reached the target.
- Found that typical chain length was 6
- Led to the famous phrase (Guare)



Back of the envelope explanation

- Ego 1
- Ego’s friends 100
- Their friends $100^2 = 10K$

$$100^5 = 10 \text{ billion} > \text{Earth's Population!}$$



Critical Property: When number of friends small compared to population, and social ties created at random probability of Ego’s friends being friends of each other is negligible

A problem...

- Random ties, however, are **not** realistic
- In reality, social networks exhibit
 - Homophily (Merton and Lazarsfeld, 1954)
 - Triadic closure (Rapoport, 1957)
 - Focal closure (Feld, 1981)
 - Spatial dependency (Festinger et al. 1950)
- Result is structure at multiple scales:
 - Clustering in network neighborhoods
 - Group affiliations
 - Communities and Organizations
 - Cities, states, and nations

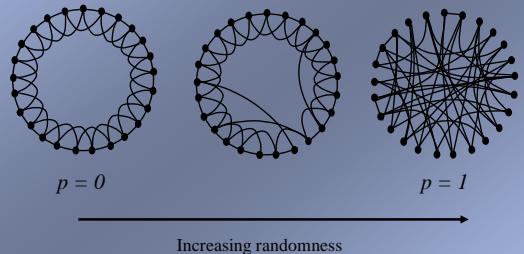
Interesting Small World Problem is therefore:

- How is it possible for Social Networks to be:
 - Very highly ordered *locally* (like social groups), and
 - Still be “small” *globally*? (like random networks)
- Problem is that **Structure** makes Analysis Hard
 - It was theoretical difficulty that led to Milgram’s experimental approach in the first place

Small World Networks

- After Milgram, not much done for 30 years
 - Theory impossible with pencil and paper
 - Experiments are hard to perform
 - Large-scale network data hard to collect
- Arrival of modern computers and the Internet enabled new approaches
- Using simulation approach, Watts and Strogatz (1998) asked: what are the conditions required for **any** network to be
 1. Locally “ordered” and
 2. Globally “small”?

Rewiring networks from Order to Randomness



Path Length (L) and Clustering (C)

$p = 0$ (Ordered)

$$L \propto \frac{n}{k}$$

$$C \approx \frac{3}{4}$$

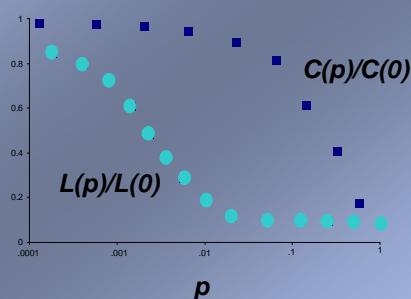
$p = 1$ (Random)

$$L \propto \frac{\ln n}{\ln k}$$

$$C \approx \frac{k}{n} \rightarrow 0$$

Intuition: the world can be
either “large and highly clustered”,
or “small and poorly clustered”,
but not “small and highly clustered”

Path Length $L(p)$ and Clustering $C(p)$ versus Random Rewiring (p) normalized by their values at $p=0$



Small-World Networks

- Main result:
 - For large N , a small **fraction** (p) of shortcuts will contract (global) **Length**, but leave (local) **Clustering** unchanged.
- Required conditions are trivial
 - Some source of “order”
 - Some source of randomness
- Conclusions:
 - Small-World Networks are generic
 - Should be widespread
 - Not confined to social networks

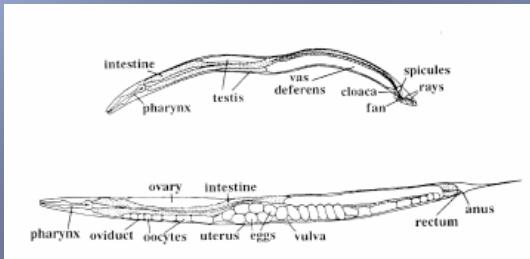
Movie Actor Graph (Aka “The Kevin Bacon Game”)



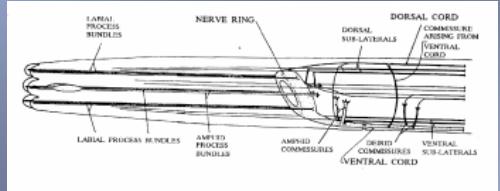
Power Transmission Grid of Western US



C. elegans



Neural network of *C. elegans*



Almost ten years later...

- We (collectively) have a good understanding of how the small world phenomenon works
- Also starting to understand other characteristics of large-scale networks
- New theories, better models, faster computers, and electronic recording all contributing to rapid scientific advance
- Result has been called “Science of Networks”
 - 2005 NAS report on Network Science
 - Many ideas borrowed from “network analysis” and graph theory but many new ideas as well

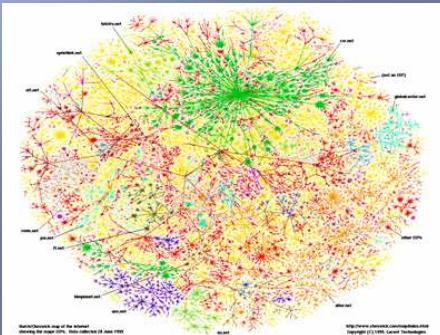


Networks are Everywhere

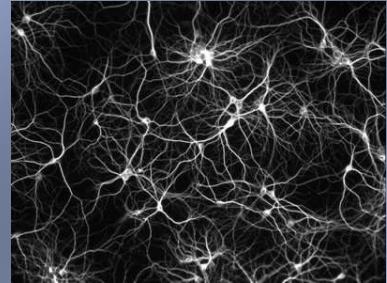
- Physical
 - Power grids, roads, airlines, Internet
- Biological
 - Neural, metabolic, genetic, ecological
- Social
 - Friendships, affiliations, sexual
- Organizational
 - Firms, markets, governments, NGO's
- Knowledge
 - Citations, words, WWW



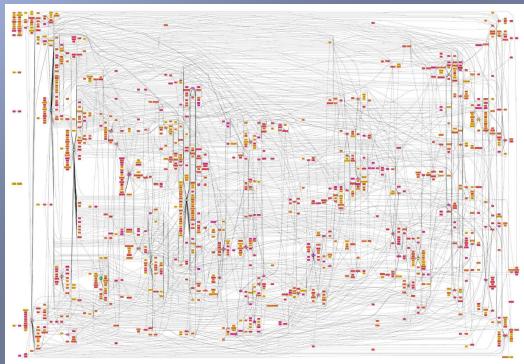
Network of Internet Domains
Hal Burch and Bill Cheswick (Lumeta Corp)



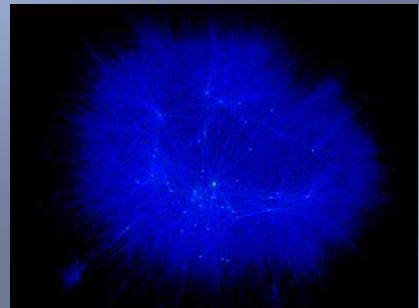
Dissociated culture of rat hippocampal neurons
Paul De Koninck Laboratory, Universite Laval (2005)



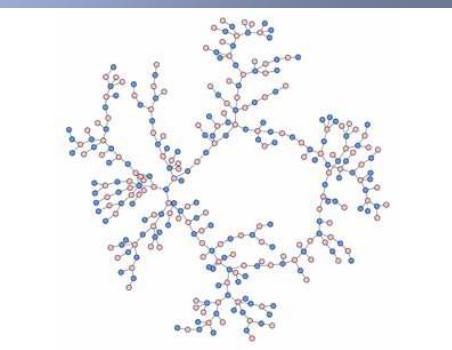
Social Network of LambdaMOO
(Charles Isbell, Michael Kearns, ATT Labs)



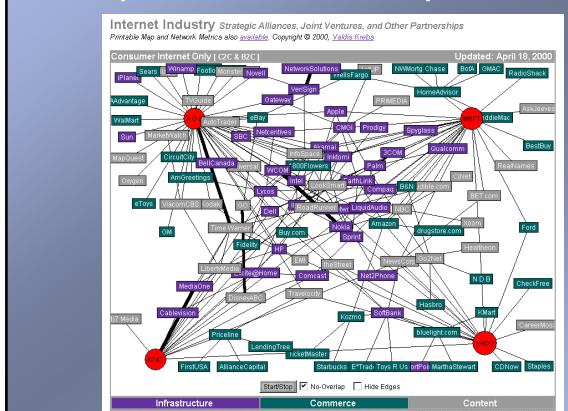
Personal Friendster network to 3 hops
(Jeffrey Heer, UC Berkeley, 2004)



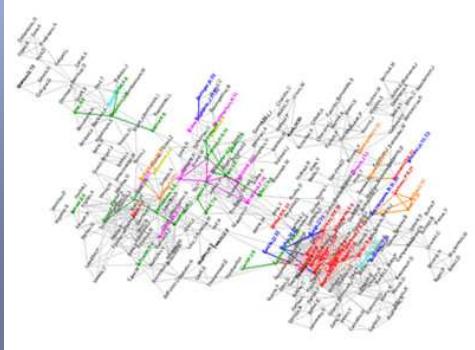
Romantic relations of high school students
(Bearman, Moody, Stovel)



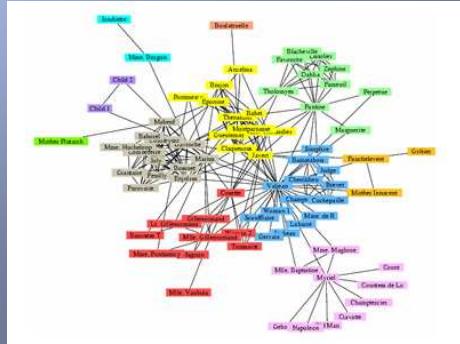
Corporate Partnerships



Citation Network of Physics Papers (Sid Redner, BU)



Interactions of Les Misérables Characters (Girvan and Newman)



But why are (social) networks interesting?

- Understand social processes
 - Rumor spreading, mob violence
 - Community formation, market dynamics, cultural change
- Infer from observed interactions:
 - Hidden / likely links
 - Organizational structure
 - Social Structure
- Control / avoid epidemics of disease
- Manage informal networks to facilitate
 - Individuals searching for resources
 - Organizations solving complex problems

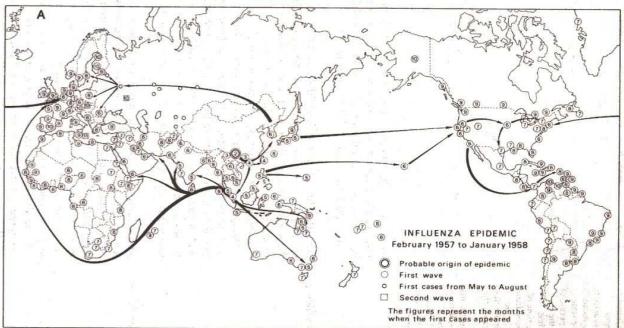


1. Networks and Epidemics

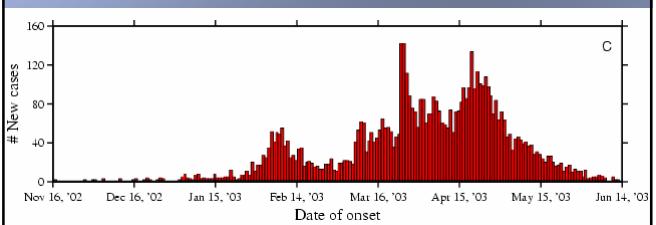
- Whenever a novel outbreak of infectious disease is announced (SARS, Avian Influenza, Ebola, etc.), one of the most pressing questions is: "How big will it get?"
- Historically, epidemic size varies dramatically
 - 1918-19 "Spanish Flu" ~ 500,000 deaths in US (20-80 Million world-wide)
 - 1957-58 "Asian Flu" ~ 70,000 deaths in US
 - 1968-69 "Hong Kong Flu" ~ 34,000 deaths in US
 - 2003 SARS Epidemic ~ 800 deaths world-wide
- All these diseases have about same R_0
- What accounts for the huge differences?



Epidemics spread in stages



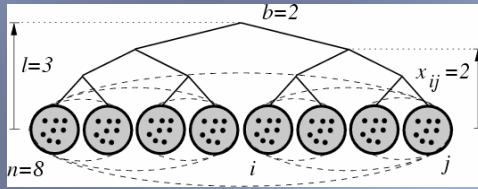
Epidemics also "Resurgent"



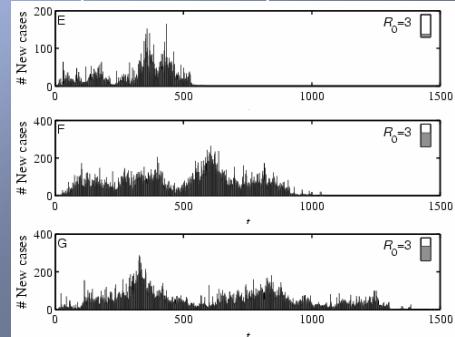
Global Daily Case Load for 2003 SARS Epidemic:
Epidemic had *several* peaks, interspersed with lulls

Importance of Network Thinking

- Large populations exhibit network structure
 - Social, sexual, infrastructure, transportation
- Large epidemics are really many small epidemics linked by networks (Watts et al. 05)
- Think of world as nested partition of groups with individuals being transported across groups



Transport of individuals on networks has dramatic implications for epidemic size



- Same disease can have very different trajectories
- Resurgence driven by “rare events” (i.e. single person)

2. Social Influence and Collective Decisions

- In cultural and other markets, “quality” hard to assess and so preferences are frequently unclear
- Too many alternatives to check anyway
 - Tens of thousands of books in a single B&N
 - Millions of songs on iTunes, Rhapsody, etc.
- Individual choices are therefore influenced by observations of others
 - Social learning
 - Conformity, coordination
 - Recent obesity study (Christakis and Fowler, 07)
- “Social influence” makes sense at individual level
 - “Ecological rationality” (Gigerenzer et al.)
- But may have counterintuitive consequences for collective (e.g. market) behavior

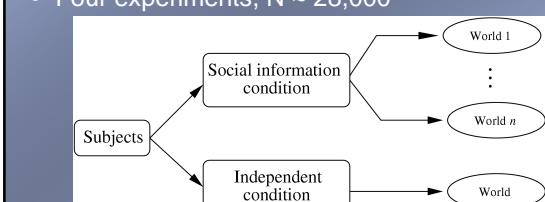


MusicLab: A web-based “artificial cultural market” (Salganik, Dodds, Watts, 06/07)



Experimental Design

- As subjects arrive, they are allocated randomly into one of two conditions
 - Independent (no download counts)
 - Social Influence (see download counts)
 - Broken into eight (8) “worlds.” can see downloads of previous participants in that world only
- Four experiments, N ~ 28,000



#	Song	# of downloads
1	HARTSFIELD: "enough is enough"	20
2	DEEP ENOUGH TO DIE: "for the sky"	17
3	THE BROKEN PROMISE: "the end in hand"	19
4	THIS NEW DAWN: "the bushes above the answer"	12
5	NOONER AT NINE: "walk away"	6
6	MORAL HAZARD: "wedge of my life"	9
7	NOT SO MUCH: "as seasons change"	27
8	SECRETARY: "keep your eyes on the ballistics"	5
9	ART OF KANY: "intensity of a breakin' breakfast"	10
10	HYDRAULIC SANDWICH: "separation anxiety"	20
11	EMBER SKY: "its upcoming winter"	25
12	SUMMER: "the dawn	13
13	RYAN ESSAMAKER: "dear, be fair"	14
14	BEEFERS: "time to see"	12
15	HALL OF FAME: "best mistakes"	19
16	GOD MOREISH: "he does what its told"	20
17	PARKER THEORY: "she said"	47
18	MISS OCTUBER: "i'm not a girl"	27
19	POST BREAK TRAGEDY: "Brenes"	14
20	FORTHADING: "i'm not a girl"	24
21	THE CALIFORNIA: "tapped in an orange peel"	20
22	52METRO: "beknown"	9
23	THE END IN HAND: "went with the crowd"	27
24	STAR CLIMBER: "tell me"	38
25	10 THE FACTLINE: "what do we gain if it don't"	31
26	A BLINDING SILENCE: "memories and mistakes"	17
27	SUMMARA: "the bushwick boogie"	25
28	SHOOTING STAR: "bushwick wock-a-v"	13
29	UP FALLS DOWN: "a bright morning star"	14
30	SUPERPOWERED: "a plan behind direction"	12
31	SILENT FILM: "all i have to say"	61
32	UNDO: "while the world passes"	12
33	UP FOR NOTHING: "i'm sight of"	13
34	SILVERFOOT: "i'm not a girl"	17
35	STRANGER: "one drop"	10
36	FAR FROM KNOWN: "i'm not a girl"	18
37	DANTE: "box myself"	17
38	UNKNOWN CITIZENS: "falling over"	34
39	BY NOVEMBER: "i'm not a girl"	20
40	LITTLE THOUGHT: "with me lack"	16
41	DRAWN IN THE SKY: "say the ride"	12
42	SILSILUS: "tan of the city"	22
43	SILENT FILM: "i'm not a girl"	14
44	EVEN GOLD: "about doorway i"	10
45	BERNIE PERRY OF A DOUBT: "i'm not a girl"	38
46	SHIPWRECK UNION: "out of the woods"	16

Results:

- Individuals are influenced by their observations of the choices of others
 - The stronger the social signal, the more they are influenced
- Collective decisions are also influenced
 - The popular songs are more popular (and unpopular songs are less popular)
 - However, which particular songs become the popular ones becomes harder to predict
- The paradox of social influence is that
 - As individuals have more information...
 - collective choice reveals *less* about preferences
 - Analogous to “information cascades”



Cascades on Networks

- Music Lab study wasn't a real “network”
 - Rather, individuals arrived in sequence
- Conducting network experiments on this scale still too hard
 - Largest experiments so far ~ 40 individuals (Kearns et al, Science, 2006)
- Can do mathematical/simulation studies
- Recent work on “threshold” models of adoption show
 - Large “cascades” can occur only when influence network is neither too sparse nor too dense
 - “Special” individuals (opinion leaders, influentials) are less important than intuition suggests



3. “Networks” and “Networking”

- Networking is an old problem in sociology
 - Doormen in New York get jobs through social ties
 - *Very few people who apply for a job as a doorman are ever successful; yet very few actual doormen had to apply, or even were looking for a doorman job* (Bearman, 2005)
 - Many labor markets are similar (Granovetter)
 - Service providers (lawyers, accountants, hairdressers, doctors, real-estate brokers) (DiMaggio)
 - Even academic job markets
- Recently “social networking” has become an mini-industry
 - LinkedIn, Visible Path, and Spoke
 - FaceBook, MySpace, numerous “vertical” sites
- But there is surprisingly little known about social “networks” that can help one “network”
 - Plenty of “how to” networking advice, but it’s not about network structure



Networking as Social Search

- Key insight is that successful networking may require a *sequence* of referrals
 - It matters not only who you know, but who they know, etc...
 - Thus the network structure matters too!
- Networking can be thought of as a search problem (Kleinberg 1999)
 - Small-world problem is simple form of search
 - What is it about the structure of social networks that makes them “searchable”?
- Unlike simpler notion of connectivity, search requires more “social” view of network structure
 - Individuals interact not on a lattice, but in groups
 - Individuals, moreover, belong to different types of groups (family, professional, neighborhood, etc.) that “cut across” each other in complex ways



When does networking work? (Lee/Watts 07)

- Networking generally works when
 - Multiple effective dimensions are present
 - Homophily in each dimension is high
- Interesting that homophily appears necessary
 - Seems to work against connectivity
 - But “zeroing in” on target requires local structure
- Weak ties not necessary for networking
 - Intuition is that weak ties span longer distances, thus carry novel / useful information
 - Here, bridging is achieved by individuals “switching” across dimensions (i.e. from professional to social)



Conclusion: Networks are tricky

- Networks usually thought of as static substrates
 - Disease, information, influence
- But networks are dynamic processes themselves
 - Individuals create and discard ties over time
 - Static “web” mental model may be misleading
- Recent work (Kossinets/Watts 06/07) on an evolving network of a university community ($N \sim 40,000$) finds
 - Aggregate properties remain constant over time
 - But individual properties change dramatically
 - Homophily is product of structural constraints
- Both kinds of dynamics are important
 - Network structure and social structure co-evolve
 - “Who you know depends upon what you do, and what you do depends upon whom you know”



But recent causes for optimism

- Tremendous interdisciplinary interest in networks
 - Physicists, computer scientists, biologists, economists
- Technologies of the internet may lift some historical constraints on data collection
 - Email, VoIP, Phone, Chat
 - Online communities, social networking sites, collaborative software
 - Multi-player online games
- Observational data that can be recorded in real time for large populations, and linked with attributes, affiliations, etc.
 - Individual-level resolution
 - Controlled experiments
 - Potentially realistic virtual environments
- Quantity of data is astounding, but relevance sometimes unclear
 - What would all the data on MySpace or Second Life tell you?



Implications for Science Policy?

- As always, more research is needed
- Difference here is that social science has gone from data-poor to data-rich overnight
- Requires rethinking how we do social science
 - Problems are now too large and complex for individuals to solve alone
 - Almost all data is now generated in private sector
- Real progress will require funding and organization that reflects these changes
 - Idealizations of social science have historically invoked Physics as a model
 - Biomedical science now seems more appropriate



***Six Degrees:
The Science of A Connected Age (2003)***

Yahoo! Research

<http://research.yahoo.com/>

Collective Dynamics Group

Columbia University

<http://cdg.columbia.edu>