



E6885 Network Science Lecture 1: **Overview -- Social, Information, and Cognitive Network Analysis**

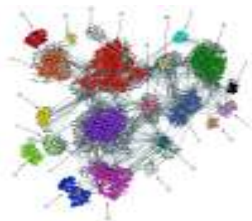
Ching-Yung Lin, Dept. of Electrical Engineering, Columbia University
September 13th, 2010



© 2010 Columbia University



Networks Everywhere

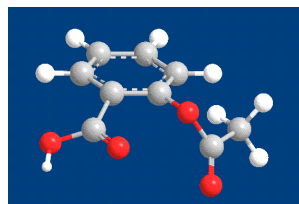


Magwene et al. Genome
Biology 2004 5:R100

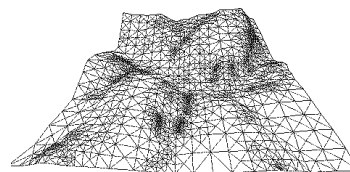
Gene Coexpression Network



Scientific Workflow



Chemical Compound



Mesh

First Reported Network Analysis

The New York Times.

Printed to Avoid One Month
Publishing, New York, N. Y.

Copyright, 1935, by The New York Times Company.
NEW YORK, THURSDAY, APRIL 6, 1935.

1700 CENTS

EMOTIONS MAPPED BY NEW GEOGRAPHY

Charts Seek to Portray the
Psychological Currents of
Human Relationships.

FIRST STUDIES EXHIBITED

Colored Lines Show Likes and
Dislikes of Individuals
and of Groups.

MANY MISFITS REVEALED

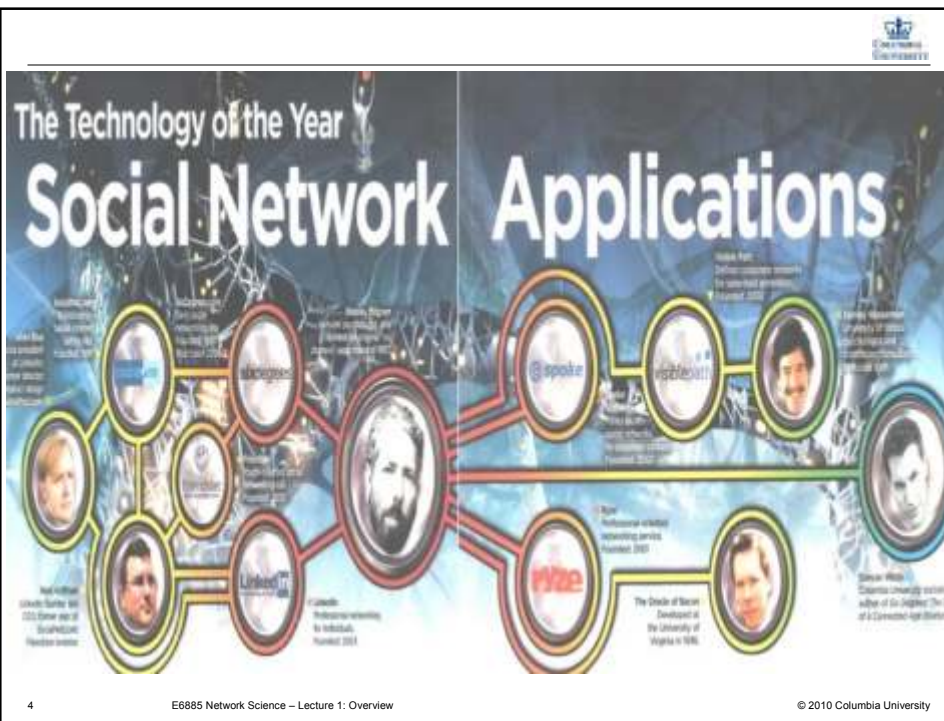
Dr. J. L. Moreno Calculates There
Are 10 to 15 Million Isolated
Individuals in Nation.

A new science, named psychological geography, which aims to chart the emotional currents, cross-currents and under-currents of human relationships in a community, was introduced here yesterday at the scientific exhibit of the Medical Society of the State of New York, which opens its 127th annual meeting here today at the Waldorf-Astoria.

The first series of maps of the new human geography were shown by Dr. Jacob L. Moreno of New York, consulting psychiatrist of the National Committee of Prisons and Prison Labor and director of research, New York State Training School for Girls, Hudson, N. Y. The maps represent studies of the forces of attraction and repulsion of individuals within a group toward one another and toward the group, as well as the attitude of the group as a whole toward its individual members, and of one group toward another group.

Emotions are represented on these psychological maps by various colored lines. Red stands for liking, black for disliking. If individual A likes B a red line with an arrow points from A to B. If B reciprocates a similar red line points from him. If he dislikes A this is indicated by a black line with an arrow pointing toward A. If B is merely indifferent the feeling is shown by a blue line.

Group of 500 Girls Studied.

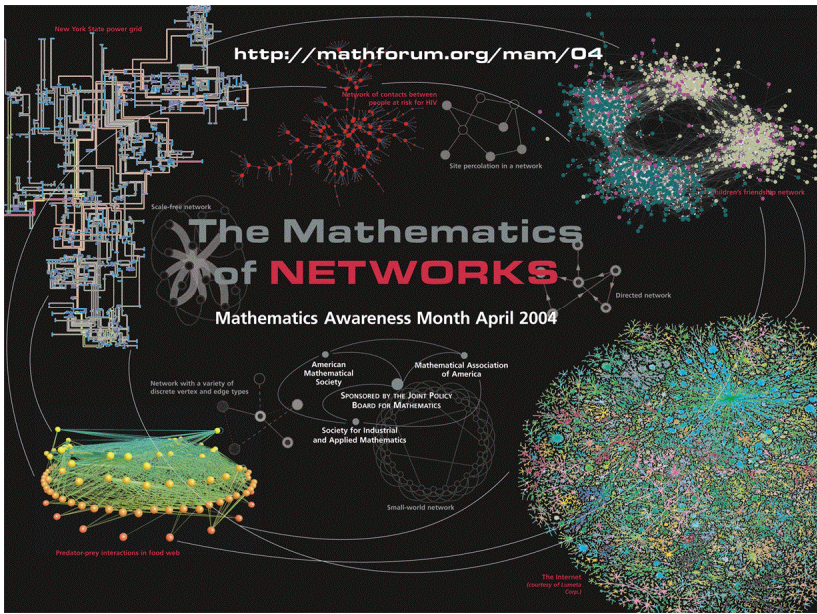




<http://mathforum.org/mam/04>

The Mathematics of NETWORKS

Mathematics Awareness Month April 2004

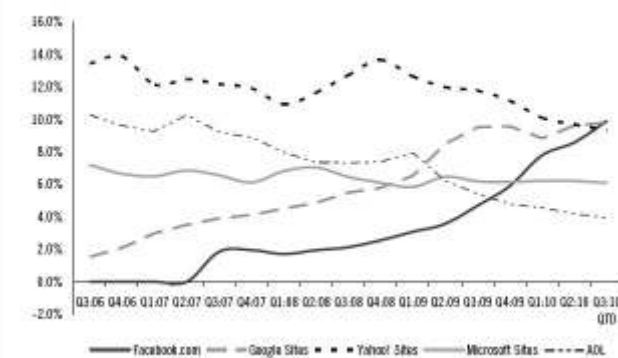


5
E6885 Network Science – Lecture 1: Overview
© 2010 Columbia University

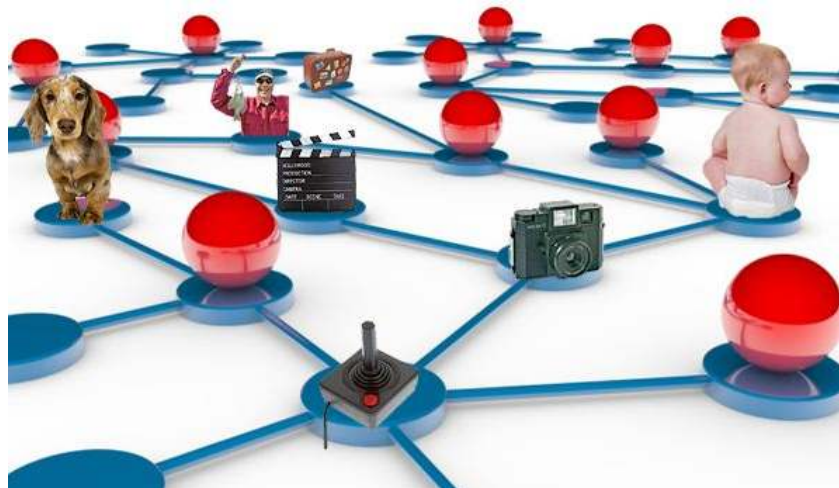
Booming of Online Social Networks

- Chomping at the bit for a [Facebook](#) I.P.O., investors are buying common stock as high as \$76 a share, giving the company a value of as much as \$33,700,000,000 based on secondary market transactions, The Financial Times – August 2010.

Figure 4. % of Time Spent vs. Total Internet (U.S.)



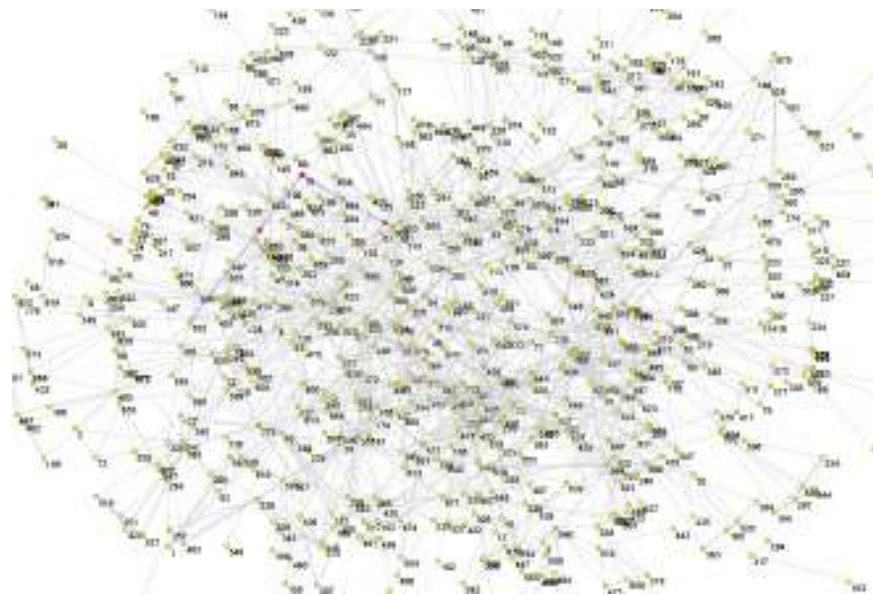
Source: comScore MediaMetrix, Citi Investment Research and Analysis



BusinessWeek Sept 2010

Characteristics of Network Data

- High-Dimensional
- Dependent
- Massive



9

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Human – a complex multimodality subject/object

- In 2005, “*Human and Social Dynamics (HSD)*” is identified as one of the five NSF key priorities among:

- Nanoscale Science and Engineering
- Biocomplexity in the Environment
- **Human and Social Dynamics**
- Mathematical Sciences
- Cyberinfrastructure



- Understanding Complexity in Natural, Social, and Build (business, economics, communications, etc) Systems -- Multi-disciplinary research relying on, and promoting advances in, **computational thinking**

Characterized by:

- Large numbers of interacting elements
- Non-linear interactions
- Dynamism, emergent behavior



10

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

The Emergence of Network Science

- Science $\Leftrightarrow$ Observable systematic empirical data
- Facility of large-scale data collection, storage and management.

The Emergence of Network Science

- Science $\Leftrightarrow$ Observable systematic empirical data
- Facility of large-scale data collection, storage and management.
- *Statistical Methodologies to combine behavior understanding, link analysis, multi-variant modeling, machine learning, graph theory, and non-parametric statistics for complex network analysis*

Contributions made by Physicists and Computer Scientists, have greatly expanded the discipline over the past 15 years.

Multi-disciplinary Research Issues

- Formation of Network
 - Communications
 - Information
 - People
 - Companies / Organizations
 - Nations
- Network Data Collection
- Network Science Infrastructure
- Network Applications
- Network Visualization
- Network Sampling, Indexing and Compression
- Network Flow
- Network Evolution and Dynamics
- Network Impact
- Cognitive Networks

13

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Multi-disciplinary Research Issues

- Formation of Network
 - Communications ← Electrical Engineering
 - Information ← Computer Science
 - People ← Sociology, Public Health
 - Companies / Organizations ← Economics, Management, Politics
 - Nations ← International Relationships, History
- Network Data Collection ← Law
- Network Science Infrastructure
- Network Applications
- Network Visualization ← Arts, Math
- Network Sampling, Indexing and Compression ← Math
- Network Flow ← Physics
- Network Evolution and Dynamics
- Network Impact
- Cognitive Networks ← Bio, Cognition, Behavior Science

14

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Network Science Consortium (2009 – 2014 or 2019)

- U.S. Army is investing \$172 million in 10 years to fund 4 *Network Science* academic research centers:

- Communication Network Academic Research Center (CNARC)
- Social and Cognitive Network Academic Research Center (SCNARC)
- Information Network Academic Research Center (INARC)
- Interdisciplinary Research Center (IRC)

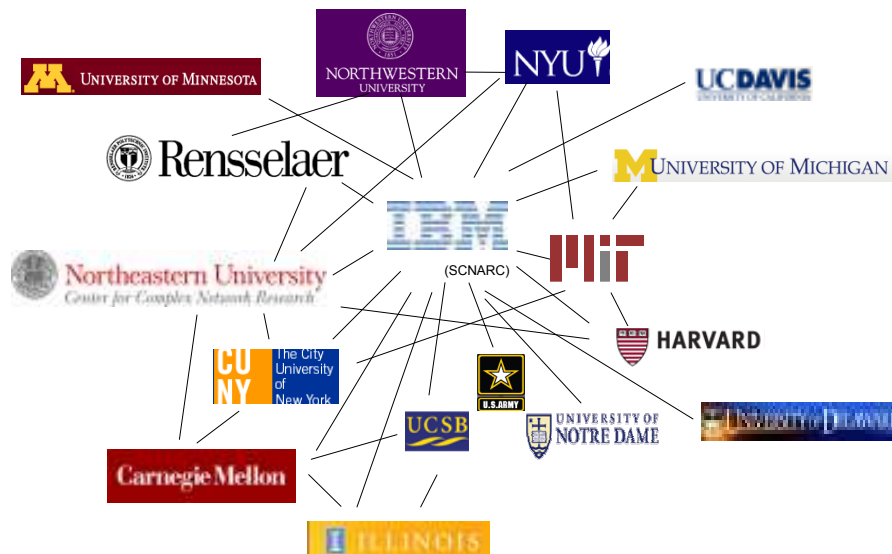
~ 100 Professor/Researcher Principle Investigators + 250 RAs, Postdocs

Objectives:

- Improve Decision Making
- Develop measures, metrics and models that describe and predict human-network interaction and exchange within & across network layers
- Develop and validate theory of human-system interaction in network-centric environments
- Explore techniques for dynamic, flexible, adaptive, and adaptable interaction

1. Charactering and Measuring Networks
2. Understanding Networks for Analysis
3. Controlling and Managing Networks
4. Using Networks

Collaborative Researches for Social and Cognitive Network Science



About this Course

Course Structure

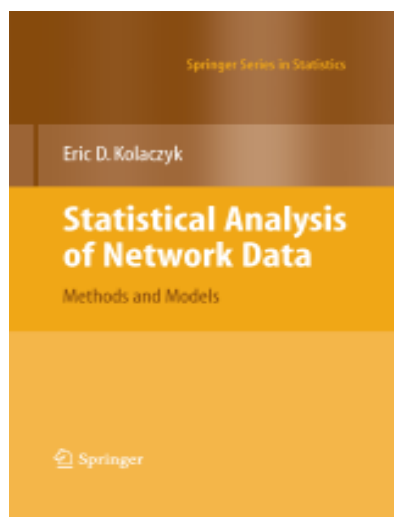
Class Date	Class Number	Topics Covered
09/13/10	1	Overview – Social, Information, and Cognitive Network Analysis
09/20/10	2	Network Representations and Characteristics
09/27/10	3	Network Partitioning and Clustering
10/04/10	4	Network Visualization
10/11/10	5	Network Sampling and Estimation
10/18/10	6	Network Models
10/25/10	7	Network Topology Inference
11/08/10	8	Dynamic Networks
11/15/10	9	Social Influence and Info Diffusion in Networks
11/22/10	10	Information and Knowledge Networks
11/29/10	11	Privacy, Security, and Economy Issues in Networks
12/06/10	12	Behavior Understanding and Cognitive Networks
12/13/10	13	Large-Scale Network Processing System
12/20/10	14	Final Project Presentation

Course Grading

- 3 Homeworks: 50%
- Final Project: 50%

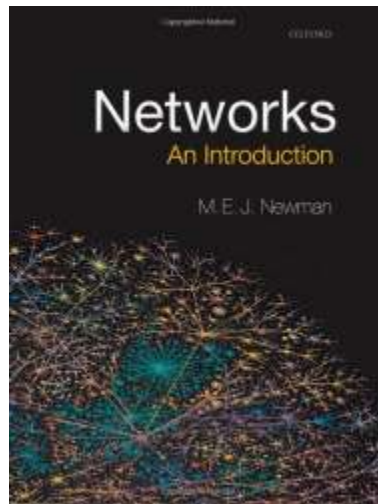
Course Textbook

- E. D. Kolaczyk, “Statistical Analysis of Network Data”, Springer, 2009



Reference Book

- M. E. J. Newman, "Networks – An Introduction", Oxford Univ Press, 2010



Other Issues

- Monday 6:40pm – 7:10pm (Mudd 1312)
- TA / Grader: TBA
- Contact: cylin@ee.columbia.edu
- Telephone: 914-784-7822
- Twitter: CUNetSci

An Example of Network Science Research: Network Analysis and Productivity

(Wu, Lin, Aral, Byrnejolfsson, 2009)



What Drives Information Worker Productivity?

- Human Capital
 - Intelligence
 - Experience
 - Education

- Financial/Physical Capital
 - Equipment
 - Access to funding sources

- Social Capital
 - How does social capital influence productivity?
 - What type of social capital is important at various work settings?

Some Components of Social Capital?

- Network Topology
 - Focus on network structures.
 - Dominant in past literature
- Network Content
 - Attributes of nodes and links inside the network
 - Information transmitted within the network
- Network Formation
 - Network dynamics
 - Speed of network formation or disintegration

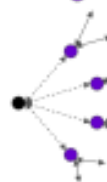
Network Topology and Worker Productivity

- Topological point of views
 - What type of network structure is beneficial?



Cohesive Network

- Trust
- Absorptive capacity
- Precision, Reliability



Structurally Diverse Network

- Brokering position
- Access to many pools of diverse, novel information

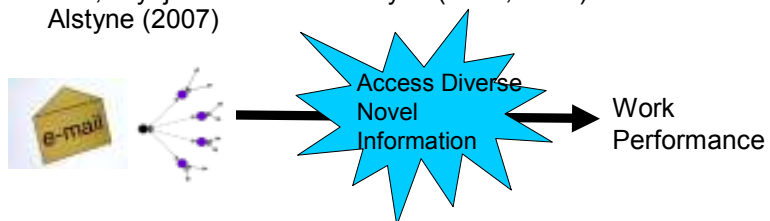


What type of network structure is most beneficial in a electronic network for consultants?

- Importance of Direct Contacts?
- Importance of Indirect Contacts?
- Constrained vs. unconstrained?

Prior Researches –Information, Networks & Productivity

- Network structure is associated with productivity and performance—focus on topology
 - Wage, promotion and job placement, innovation (Burt 1992, 2004)
 - Productivity of R&D teams (Reagan & Zuckerman 2001)
- Access to diverse and novel information impacts work performance
 - Aral, Brynjolfsson & Van Alstyne (2006, 2007) and Aral & Van Alstyne (2007)



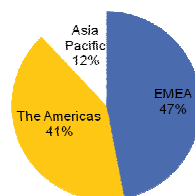
Information Content of Social Networks

- The scope of prior studies are limited
 - Small sample
 - Limited context
- Information content needs to be explored
 - Type of knowledge and information exchanged inside a network
 - Timing of communications
 - Attributes of people inside the network
 - Characteristics of ties
- **Biggest Challenge: Very difficult to observe and gather long-term large-scale social network and financial performance data**

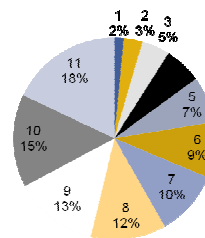
We studied the impact of network characteristics towards revenue in the personal and project levels on consultants

- We studied the network properties of 2600+ anonymized IBM Global Business Services consultants over 2 years (2006-2008) in a context of 7,500 SB volunteer network or over 300K+ inferred network* of the company. *: lower network accuracy
- Overall observation of 10k consulting projects.
- We also conducted 1-hour interviews with 15 consultants from Bands 6-11 in Canada & US to gain insights on how consultants work.
- All data are anonymized

Business by Region



Bands



29

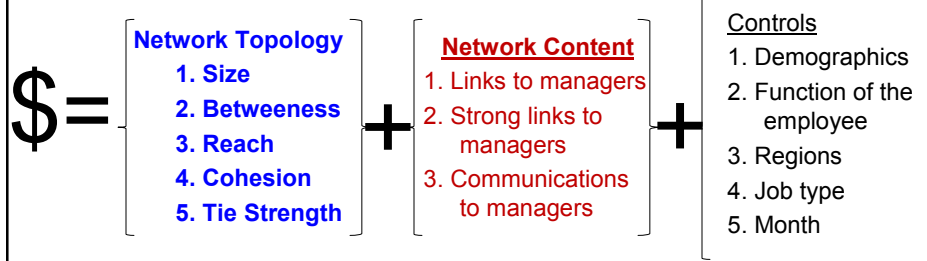
E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Linking Performance to Social Networks

Statistical Specifications:

$$Y_{it} = \alpha + \beta_1 \text{hours}_{it} + \beta_2 \text{Network Variables}_{it} + \beta_3 \text{Network Content} + \text{Controls}_{it} + \varepsilon$$

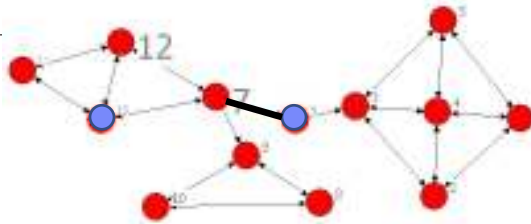


30

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Network Topology Measures



Direct Contacts	Size(7) = 4 Size(12)= 3	+ No information distortion - High maintenance cost	Network size → strong work performance (?)
Indirect Contacts	Btw(7)= 33 Btw(12)=6 3steps(7) =11 3steps(12)=8	+ Access diverse information - Information distortion	Btw-centrality → Strong work performance (?) 3-step Reach → Strong work performance (?)
Structural Diversity	Div(7)=.53 Div(12)=0.16	+Transfer complex knowledge - Access diverse knowledge	Diversity→ Strong work performance (?)

31

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Network Content

$$Y_{it} = \alpha + \beta_1 \text{hours}_{it} + \beta_2 \text{Network Variables}_{it} + \beta_3 \text{Network Content} + \text{Controls}_{it} + \varepsilon$$

- Attributes of direct contacts
 - Number of managers
 - Number of divisions
- Attributes of tie strength
 - Ego Strength as measured by communication frequency
 - Normalized between 0 and 1
 - Strong links: defined to have ego strength > 0.8*

* We tried different specifications of strong links. They do not qualitatively change our results

32

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Types of Analyses

- People-Level Social Network
 - (1) Long-term networks
 - Network and Financial observations are based on all data collected in the entire time-span.
 - Cross-Sectional Model
 - (2) Short-term dynamic networks
 - Network and Financial observations are based on month-by-month data
 - Look at the changes through time and the difference
 - Fixed-Effect Model
- (3) Project-Level Network
 - In the network graph,
 - each node represents a project
 - each edge is an aggregated relationships between two project nodes.

Preliminary Finding—Personal Social Network vs Revenues

Table 3: Person-level Email Networks	
Dependent Var.	Personal Revenues in a month
Controls: Average Project Complexity, Line of Business, Months, Regions, Job level	
Hours	\$119.6*** (2.307)
Size	\$74.07*** (26.38)
Betweenness	-\$348512*** (92092)
Num ppl reachable in 3steps	\$0.163*** (0.0168)
Num of stronglinks	\$-7.920*** (0.947)
Constraint	\$-758.5** (119.6)
Direct Links to managers	\$1074* (643.6)
isManager	\$-1322 (1040)
Gender	\$44.56 (165.4)
Observations	6091
R-squared	0.751
*** p<0.0001 ** p<0.05 *p<0.1	

Production Function Framework

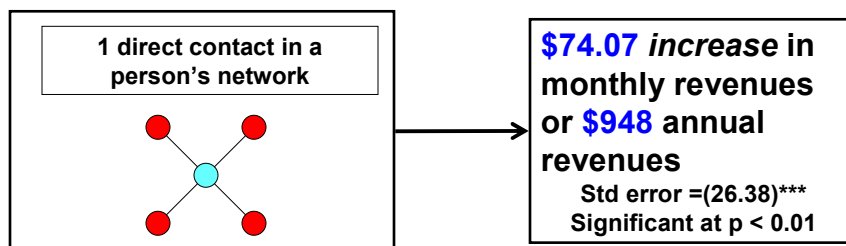
1. **Dependent variable: revenues** generated for a person in a month
2. **Network: each node is a person, each link is the total communications between 2 people.**

Productivity effect from network variables

- An additional person in network size ~ \$74 revenue per month
- Each person that can be reached in 3 steps ~ \$0.163 in revenue per month
- A link to manager ~ \$1074 in revenue per month
- 1 standard deviation of network diversity (1 - constraint) ~ \$758
- 1 standard deviation of btw ~ -\$300K
- 1 strong link ~ -\$7.9 per month

Relationship Between Network Topology and Productivity

- Network size is positively correlated with performance.
 - Each person in your email address book at work is associated with \$948 dollars in annual revenue.



Observations from Personal Social Networks vs. Revenue

- Structural Diverse networks with abundance of structural holes are associated with higher performance.
 - *Having diverse friends helps.*
- Betweenness is negatively correlated.
 - *Being a bridge between a lot of people is not helpful.*
- Network reach are highly correlated.
 - *The number of people reachable in 3 steps is positively correlated with higher performance.*
- Having too many strong links — the same set of people one communicates frequently is negatively correlated with performance.
 - *Perhaps frequent communication to the same person may imply redundant information exchange.*

Dynamic Network Analysis

- Network topology and financial data are month by month

- Benefit of such analysis:

- Observing 'changes' can eliminate unobserved fixed characteristics in the model, which gives a good chance to increase the model accuracy

- Drawback of such analysis:

- Data amount are then much smaller and then are more sparse.

Diverse Network • • • • • Performance

An 'enthusiastic' person

example of unobserved characteristic

$$Y_{it} = \alpha + \beta_1 \text{hours}_{it} + \beta_2 \text{Network Variables}_{it} + \beta_3 \text{Human Capital of Social Contacts}_{it} + \text{Controls}_{it} + \text{unobservables}_{it} + \varepsilon$$

$$Y_{it+1} = \alpha + \beta_1 \text{hours}_{it+1} + \beta_2 \text{Network Variables}_{it+1} + \beta_3 \text{Human Capital of Social Contacts}_{it+1} + \text{Controls}_{it+1} + \text{unobservables}_{it+1} + \varepsilon$$

$$\Delta Y_{it} = \alpha + \Delta \beta_1 \text{hours}_{it} + \beta_2 \Delta \text{Network Variables}_{it} + \beta_3 \Delta \text{Human Capital of Social Contacts}_{it} + \text{unobservables}_{it} - \text{Controls}_{it} + \varepsilon$$

Fixed-Effect Model

Dependent Var.	Monthly Personal Revenues
	Fixed Effect
Betweenness Centrality	513.35% ** (217.83)
Constraint	-276.64% *** (113.88)
Size	\$17.82 (20.70)
Number of strong links	\$56.10 (48.43)
Number of managers in network	-\$98.48* (53.49)
Number of strong links to managers	\$588.2* (389.4)
Total communication to managers	-\$6.777 (4.880)
Divisions in network	-\$196.6 (128.7)
Observations	5527
R-squared	0.81

Personal Social Networks and Productivity based on Dynamic Network Analysis

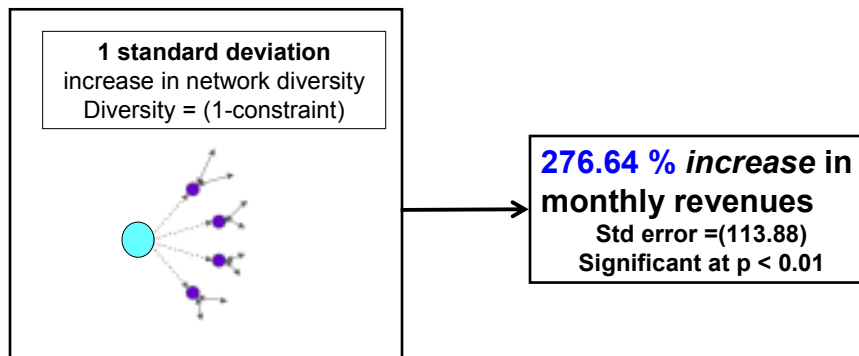
Production Function Framework

1. Dependent variable: revenues generated for a person in a month
2. Network: each node is a person, each link is the total communications between 2 people.
3. Fixed-effects model: focus on changes of network structure and revenues.

*p<.1, **p<.05, ***p<.001. Huber-white robust standard errors are shown in parentheses

Relationship Between Network Topology and Productivity

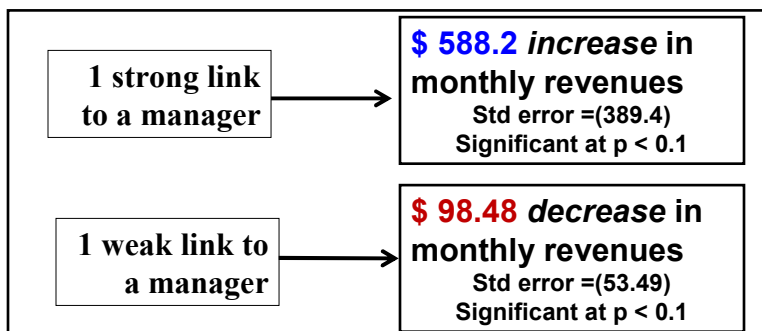
- Structural diverse networks with abundance of structural holes are associated with higher performance.
 - When friends of your friends are not friends of each other or belong to the same social group.



Content of the Network and Productivity

Having *strong links* to *managers* is correlated with performance

- strong links to anyone else is not correlated with performance



Selectively cultivate a few relationships with management.

Content of the Network and Productivity

We do not observe a clear relationship between performance to personal

1. access to different divisions
2. access to different geographical locations
3. gender distribution

- Diversifying personal direct network (Ego network) does not show significant contribution to performance.

- Diversifying personal overall network shows significant contribution to performance.

→ For privacy reason, you cannot look beyond your direct network. However, a social network analysis tool that knows overall network can help diversifying your network.

41

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Dependent Var.	Monthly Project Revenues
	212.11** (69.5)
Network Size	304.36*** (100.3)
Betweenness	77.42*** (20.13)
Betweenness range	146.81 (158.85)
Constraint	6395*** (2267)
Direct links to mgrs outside of the project	2733.9*** (537.5)
# Managers in project	-682.02*** (215.3)
(# Managers in project) ^2	4533*** (1441.00)
Gender	-4374*** (1416)
Gender^2	-513.41*** (128.55)
Number of divisions project can access	8018
Observations	0.913
R-squared	

Project Social Networks

Production Function Framework

- **Linear regression**
- Dependent variable: revenues generated for a project in a month
- Network: each node is a project. Each link is the total communication between two projects.
- Total of 8018 observations.

Additional Controls

- Number of people in projects
- Project characteristics: complexity, line of business, regions, month, the highest job level in the project.

*p<.1, **p<.05, ***p<.001. Huber-white robust standard errors are shown in parentheses

42

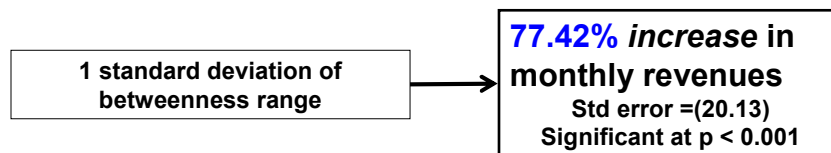
E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

What is Driving Project Centrality?

- Are **all team members** at the center of information highway good?
- Or
- Just a few** is better?

- We can test this using the range of individual betweenness centralities in a team. (Btw range = max – min of individual btw)



- Optimal team composition is not to have all super stars, but a complementary team members with a few information keepers who are in the middle of information highway.

43

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

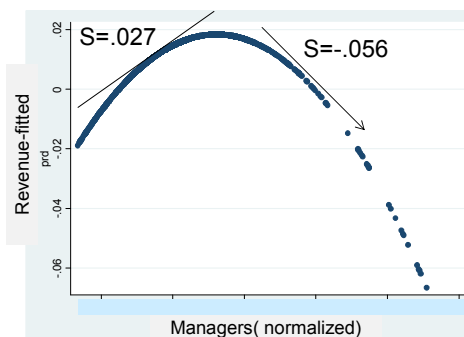
Project Team Composition—Managers

The number of managers in a project exhibit an inverted-U shaped curve.

- Having managers in a project is correlated with team performance initially.
- Too many managers in a project is negatively associated with team performance.

$$revenue = \alpha + \beta_1 \cdot mgr + \beta_2 \cdot mgr^2 + \gamma_1 \cdot otherfactor_1 + \dots + \gamma_k \cdot otherfactor_k + \varepsilon$$

# Managers in project	β_1	2733.9*** (537.5)
(# Managers in project) ^2	β_2	-682.02*** (215.3)



44

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Preliminary Result of Another Study: Pre-Sales Performance

Deal hub focuses on the pre-sales productivity instead of consultants. People working on presales are responsible for bring the contracts into the firm. We studied anonymized data of 532 pre-sales.

- Network characteristics of productive pre-sales and consultants are different.
 - Strong ties are *positively* correlated with performance for presales.
 - Building a strong relationship with clients through *repeated and long term interactions* may be the key for presales.
 - Strong ties are shown to *negatively* associate with performance for consultants.
 - Network diversity may be more important for consultants as seeking and participating in projects of *wide range* may be valued more.

Summary of Key Findings

- Person and Project Networks
 - Structural diversity is positively correlated with performance
 - Strong connections with managers are correlated with better work performance
 - Demographic diversity does not have a clear relationship with performance
- Project Networks
 - Having just a few information keepers who have high betweenness centrality is positively associated with performance.
 - Managers in a project exhibit a inversed-U relationship to team performance

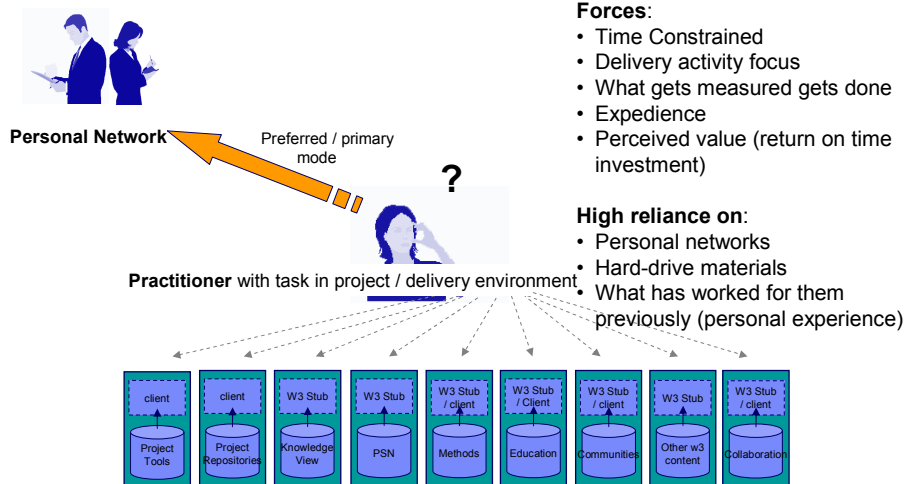
An Example of Network Science Applications

Unlocking the expertise that resides among my friends, or my friends' friends...

- “If I know who among our 400,000 IBM colleagues knows the answer of this critical question customer is asking, a million dollar deal signing may be only one phone call away...” -- *an IBM CHQ Sr. VP in 2003, showing how desperately IBM needs a novel expertise location tool...*



Gartner 2003 – Personal Network preferred source for information

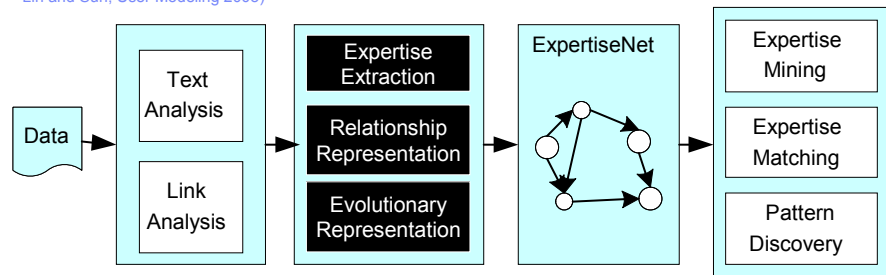


How to Unlock the power of our existing networks -- automatically?

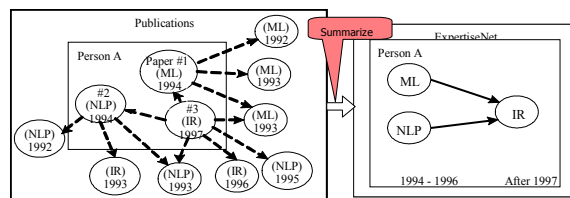
Where does knowledge reside?

Emails, Instant Messages, Calendars, Databases, Wikis, Blogs, Microblogging..

ExpertiseNet: Modeling Person's Expertise Evolution as Dynamic Graphs (Song, Tseng, Lin and Sun, User Modeling 2005)



Build graphical representations of a researcher's expertise based on her/his publications and citation networks



SmallBlue asset: Large-Scale People Modeling and Social Network Analysis

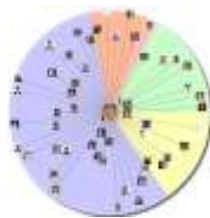


Live Data, Live Production System

20,000,000 e-mails & SameTime messages
 1,000,000 Learning click data
 14,000,000 KnowledgeView, SalesOne, ..., access data
 1,000,000 Lotus Connections (blogs, file sharing, bookmark) data
 200,000 people's consulting financial databases
 400,000 organization/demographic data
 100,000 intranet w3 searches per day



SmallBlue unlocks the Power of Business Networks and Protects Privacy



Ego: Show my personal network evolution and social capital

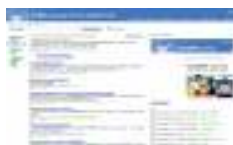


Whisper: Social Network enabled personalized live recommender..



Expertise: Search for people who know "xyz" in my networks..

Social Network Analysis & Visualization,
Expertise Mining, and Multi-Channel Human
Network/Behavior Analysis



Synergy:
Personalized Search



Net: See how experts or community connect

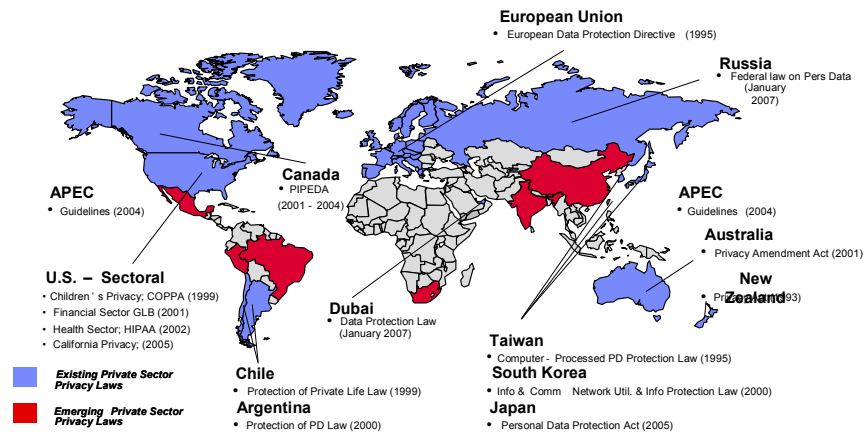


Reach: helps me understand this person, my formal and informal paths to Reach him..



Productivity: Social Network Analysis Service helps company understand how to enhance productivity.

Global Privacy Laws



55

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

SmallBlue Applications (I): Find knowledgeable colleagues

- E.g.: Search for the most knowledgeable colleagues within my 3-degree network for who knows 'healthcare'. (or within a country, a division, a job role, or any group/community)

W3 SmallBlue Suite

Home | Find | Search | List | Info | Admin

Search for (subject keywords):

Search results: 1-10 of 10 results found for 'healthcare' (10 results found)

1. Susan (Healthcare)
Global Business Services
Associate Partner, Healthcare Integration
Other Consultant
Last modified: 10/10/2009 10:10:10 AM

2. Susan (Healthcare)
Global Business Services
Associate Partner, Healthcare Integration
Other Consultant
Last modified: 10/10/2009 10:10:10 AM

3. Susan (Healthcare)
Global Business Services
Associate Partner, Healthcare Integration
Other Consultant
Last modified: 10/10/2009 10:10:10 AM

4. Susan (Healthcare)
Global Business Services
Associate Partner, Healthcare Integration
Other Consultant
Last modified: 10/10/2009 10:10:10 AM

5. Susan (Healthcare)
Global Business Services
Associate Partner, Healthcare Integration
Other Consultant
Last modified: 10/10/2009 10:10:10 AM

6. Susan (Healthcare)
Global Business Services
Associate Partner, Healthcare Integration
Other Consultant
Last modified: 10/10/2009 10:10:10 AM

7. Susan (Healthcare)
Global Business Services
Associate Partner, Healthcare Integration
Other Consultant
Last modified: 10/10/2009 10:10:10 AM

8. Susan (Healthcare)
Global Business Services
Associate Partner, Healthcare Integration
Other Consultant
Last modified: 10/10/2009 10:10:10 AM

9. Susan (Healthcare)
Global Business Services
Associate Partner, Healthcare Integration
Other Consultant
Last modified: 10/10/2009 10:10:10 AM

10. Susan (Healthcare)
Global Business Services
Associate Partner, Healthcare Integration
Other Consultant
Last modified: 10/10/2009 10:10:10 AM

My shortest path to Susan

As a user, you can only see their public information. Private info is used internally to rank expertise but private data can never be exposed.

Click a name to see their profile (SmallBlue Reach)

56

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

SmallBlue Applications (II): Reach – social dashboards

- Is Tom a right person to me?

His official job role, title, contact info

His self-described expertise

His public communities

The public interest groups he is in

His blogs, forum, postings..

My various paths to Tom. SmallBlue can show the paths to any colleagues up to 6-degree away

57

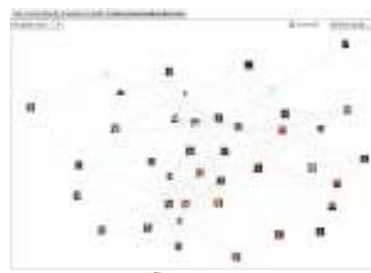
E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

SmallBlue Applications (II): Reach (cont'd)

- Is Tom a right person to me?

Formal hierarchy path, if I really want to go through this 13-degree path...



The network of people we know in-between us.

58

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

SmallBlue Applications (III): Net – corporate social network analysis

- How are company's top healthcare experts link with each other? Who are the key bridges? Who have the most connections? How do these experts cluster?



SmallBlue Applications (III): Net (cont'd)



Healthcare experts in the world



Healthcare experts in the U.S.



Connections between different divisions



Key social bridges

SmallBlue Applications (IV): Ego – personal social network capital management (an application only visible to the user himself)

- What is a friend's social capital to me?



61

E6885 Network Science – Lecture 1: Overview

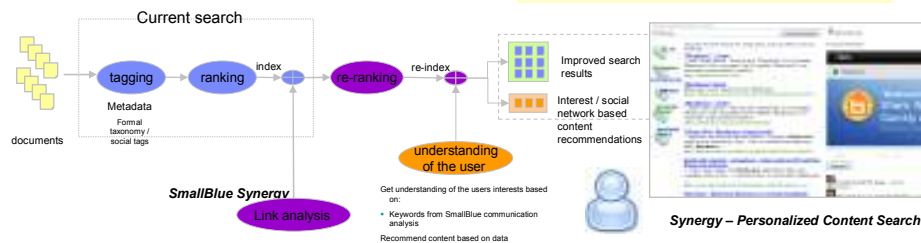
© 2010 Columbia University

Personalized Content Recommendation and Search

- What your friends know become what you know.. Your friends are your window to the world – *Confucius ~600 B.C.*
- Utilizing the unique large-scale weighted social network inferred by SmallBlue, personalized ranking becomes possible.
- Fusion of Recommenders: Social Filtering, Collaborative Filtering, Latent Semantic Filtering, Popularity & Freshness Filtering, etc.
- Deploy on IBM KnowledgeView, IBM Learning, and IBM TAP



46 IBMers are 1 degree away from you.
1278 IBMers are 2 degrees away from you.
14232 IBMers are 3 degrees away from you.



- SmallBlue Whisper and Synergy usages:
 - On Nov. 19, 2008: 11,108
 - On Nov. 20, 2008: 9,567

62

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

SmallBlue (Atlas) Mobile

SmallBlue Applications on Mobile Phones



SmallBlue Find Widget in Mobile



SmallBlue Whisper Widget in Mobile



Android



BlackBerry



Nokia

Show Expertise of 'SNA' inside:

- (1) IBM
- (2) My 2-degree network
- (3) Research division
- (4) Global Business Services
- (5) Any group – e.g., Distinguished Engineers

Recommend Contents from Friends within 3-degrees

63

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Several Technical Challenges

- Large-Scale Expertise ranking – Google PageRank version of People (This is not something in open sources such as Lucene)
- Real-Time Dynamic Large-Scale Social Path Graph Indexing
- Distributed Privacy-Preserving Sensor System – **what personal data shall be executed in personal control; what personal data can be sent to and processed by server..**
- Requirements for expertise/knowledge search:
 - Dirty word detector
 - Name detector
 - Sensitive query detector (dollars, confidential words, numbers, etc.)
 - Multi-language processing (Languageware components)
 - Multi-platform (Lotus Domino Server, Microsoft Exchange Server, Google/Yahoo/MSN imap server)
 - Multi-search requirements (Job roles, geography, community, any BlueGroups)
- Scalability – 7 nodes for production, 4 nodes for staging, for IBM deployment. Flexible design for different company sizes.
- Stability – transferred to Lotus Software Developers in 2007. Compliant with IBM Software development standards. Full documentations (30 architecture documents). Production Services since March 2008 – 18 months already. 713,000 usages in the last 11 months.
- Accessibility -- For Blind and Disabled Users
- System configurability and management – Web-based admin reports, Websphere configuration components..

64

E6885 Network Science – Lecture 1: Overview

© 2010 Columbia University

Questions?