

Machine Learning for Complex Social Processes

Hanna Wallach

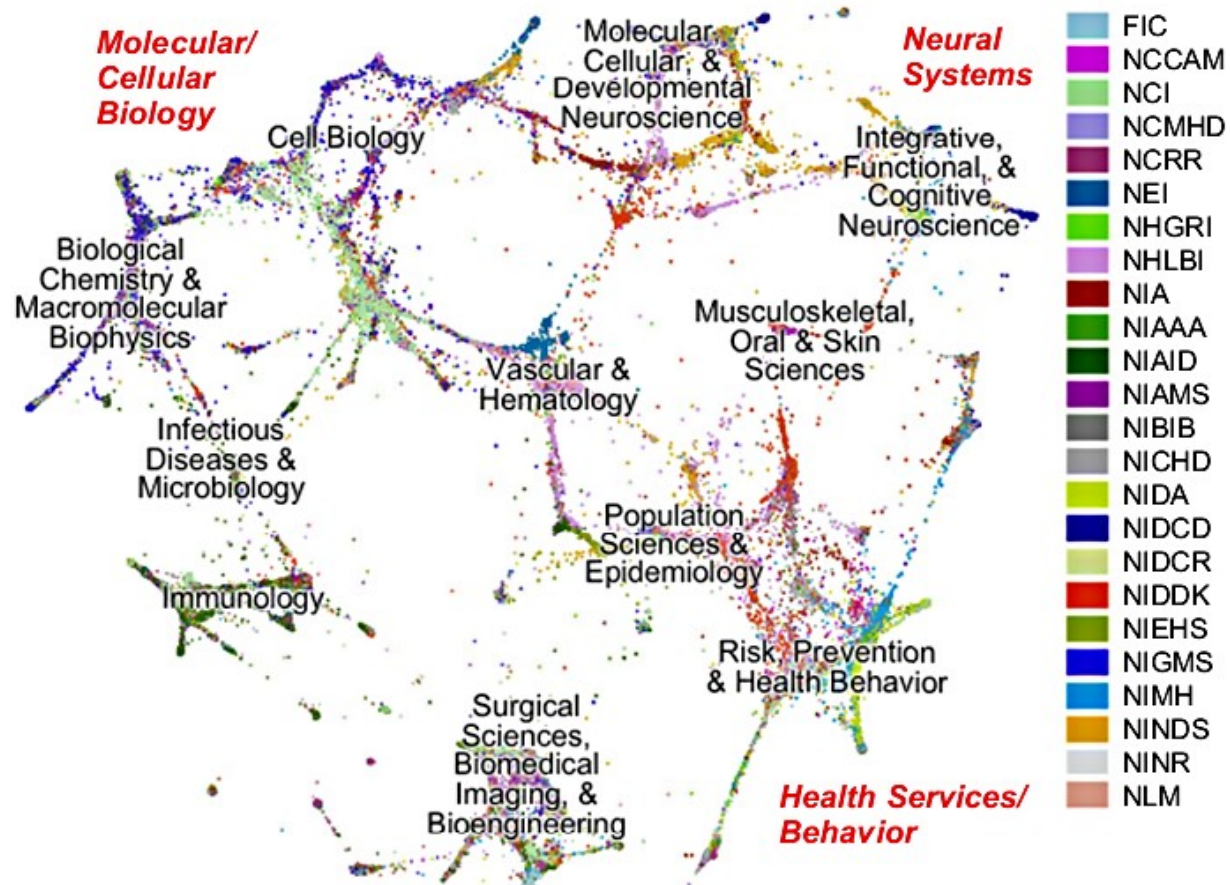
University of Massachusetts Amherst

wallach@cs.umass.edu

Complex Social Processes



National Institutes of Health



United States Patent System

(12) **United States Design Patent** (10) **Patent No.:** **US D478,999 S**
Jobs et al. (45) **Date of Patent:** **** Aug. 26, 2003**

(54) **STAIRCASE**

(75) Inventors: **Steve Jobs**, Palo Alto, CA (US); **Karl Backus**, Emeryville, CA (US); **Rosa Sheng**, Emeryville, CA (US); **Ben McDonald**, San Francisco, CA (US); **Michael Waltner**, Berkeley, CA (US); **Colleen Caulliez**, San Francisco, CA (US); **James O'Callaghan**, New York, NY (US); **Graham Coult**, London (GB); **Damian Rogan**, New York, NY (US); **Scott Nelson**, Cirencester (GB)

(73) Assignee: **Apple Computer, Inc.**, Cupertino, CA (US)

(**) Term: **14 Years**

(21) Appl. No.: **29/164,077**

(22) Filed: **Jul. 15, 2002**

(51) **LOC (7) Cl.** **25-04**

(52) **U.S. Cl.** **D25/62**

(58) **Field of Search** D25/62, 69; 52/182, 52/184, 188, 190, 191

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,022,197 A * 6/1991 Aragona 52/184
D371,581 S 7/1996 Järnros
D389,588 S 1/1998 Dunk

D398,063 S 9/1998 Kline
D399,975 S * 10/1998 Confer D25/62
D415,289 S 10/1999 Dalton
5,960,516 A 10/1999 Zoroufy et al.
D417,736 S 12/1999 Cavaness
D423,079 S 4/2000 Blount
6,059,269 A 5/2000 Ross
D428,629 S 7/2000 Cohen
D431,303 S 9/2000 Maiuccoro
6,176,027 B1 1/2001 Blount
6,205,722 B1 3/2001 Bromley et al.

* cited by examiner

Primary Examiner—Doris Clark

(74) *Attorney, Agent, or Firm*—Beyer Weaver & Thomas, LLP

(57) **CLAIM**

We claim the ornamental design for a staircase, substantially as shown and described.

DESCRIPTION

FIG. 1 is a perspective view of a staircase in accordance with the present design. The staircase has a transparent character. FIG. 2 is a front view for the staircase shown in FIG. 1. FIG. 3 is a rear view for the staircase shown in FIG. 1. FIG. 4 is a left side view for the staircase shown in FIG. 1. FIG. 5 is a right side view for the staircase shown in FIG. 1. FIG. 6 is a top view for the staircase shown in FIG. 1; and, FIG. 7 is a bottom view for the staircase shown in FIG. 1.

1 Claim, 7 Drawing Sheets

Representatives and Constituents

Pelosi Statement on Two Year Anniversary of Student Aid and Fiscal Responsibility Act

NEWS

March 30, 2012

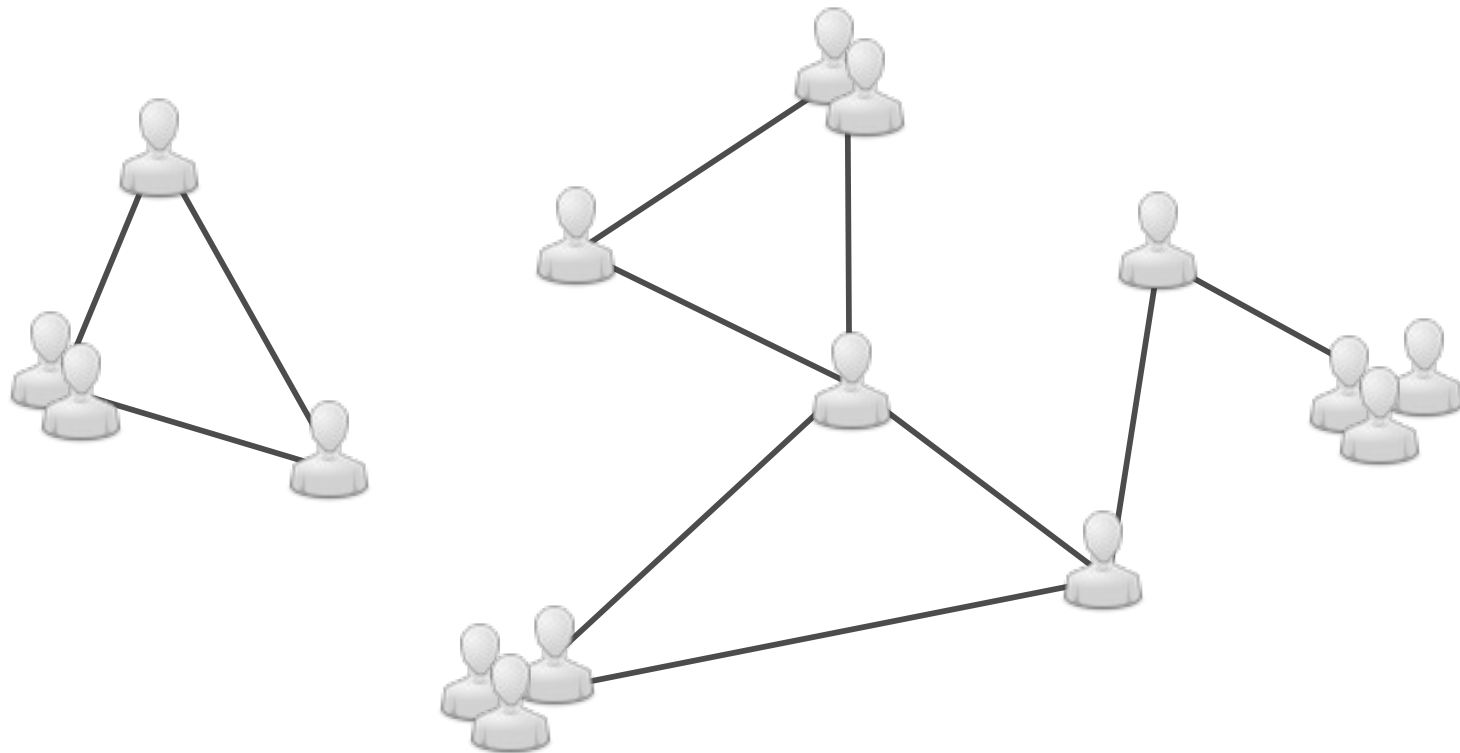
Contact: Nadeam Elshami/Drew Hammill, 202-226-7616

Washington, D.C. – Democratic Leader Nancy Pelosi released the following statement today in commemoration of the second anniversary of the Student Aid and Fiscal Responsibility Act, which represents the single largest investment in college aid in our nation's history:

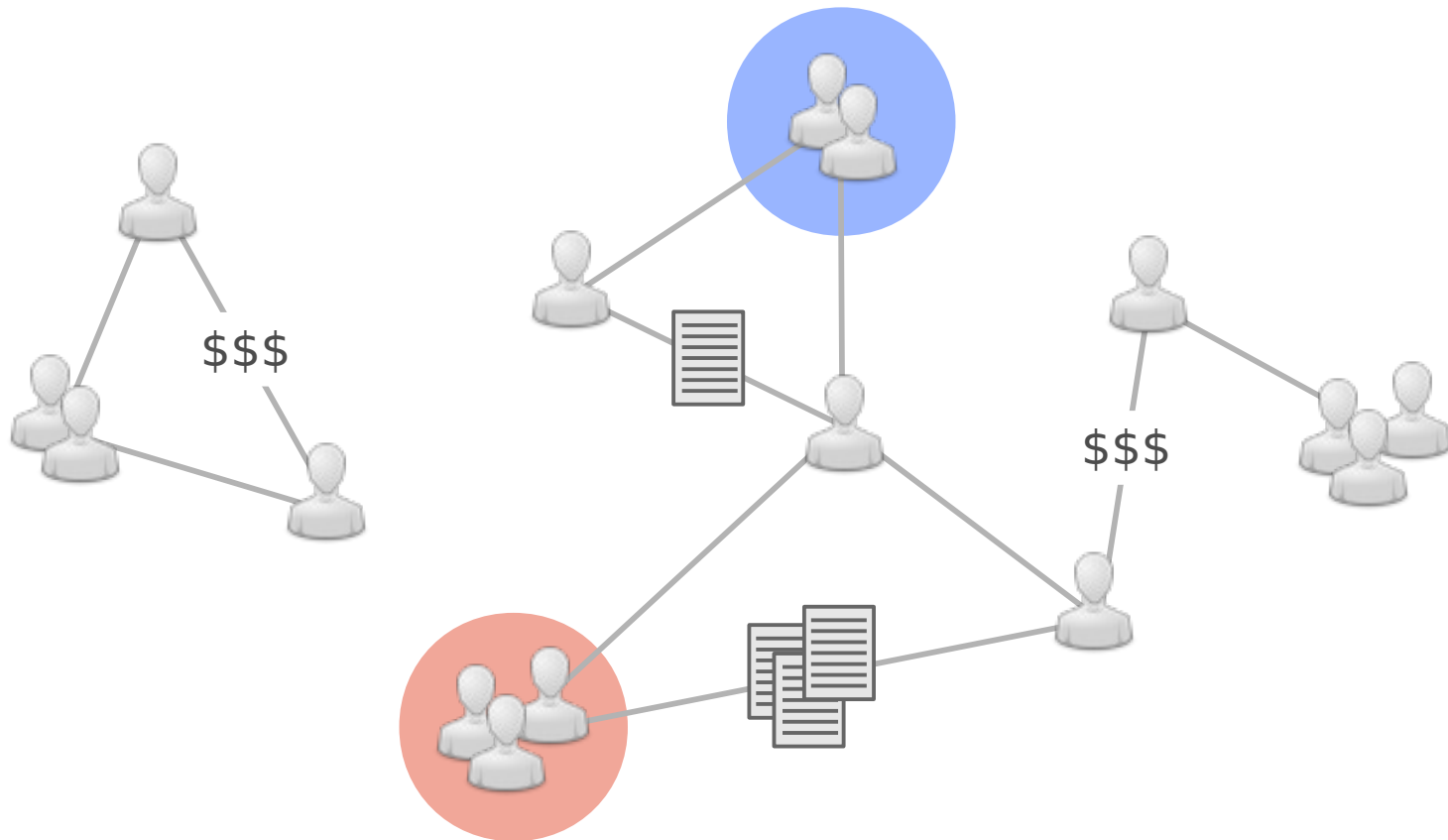
“Two years ago, Democrats were proud to lead the way in passing the single largest investment in college aid in our nation's history. With the Student Aid and Fiscal Responsibility Act, we lowered the cost of student loans, strengthened community colleges, increased the maximum Pell Grant, and invested in Historically Black Colleges and Universities and Minority Serving Institutions.

“Education is the best investment parents can make in their children, individuals can make in themselves, and a nation can make in its future. That's why the budget passed by House Republicans this week is so distressing. Instead of reigniting the American dream, it makes it more difficult for student to afford higher education: allowing interest rates on some students loans to double and cutting hundreds of thousands of students from the Pell Grant program.

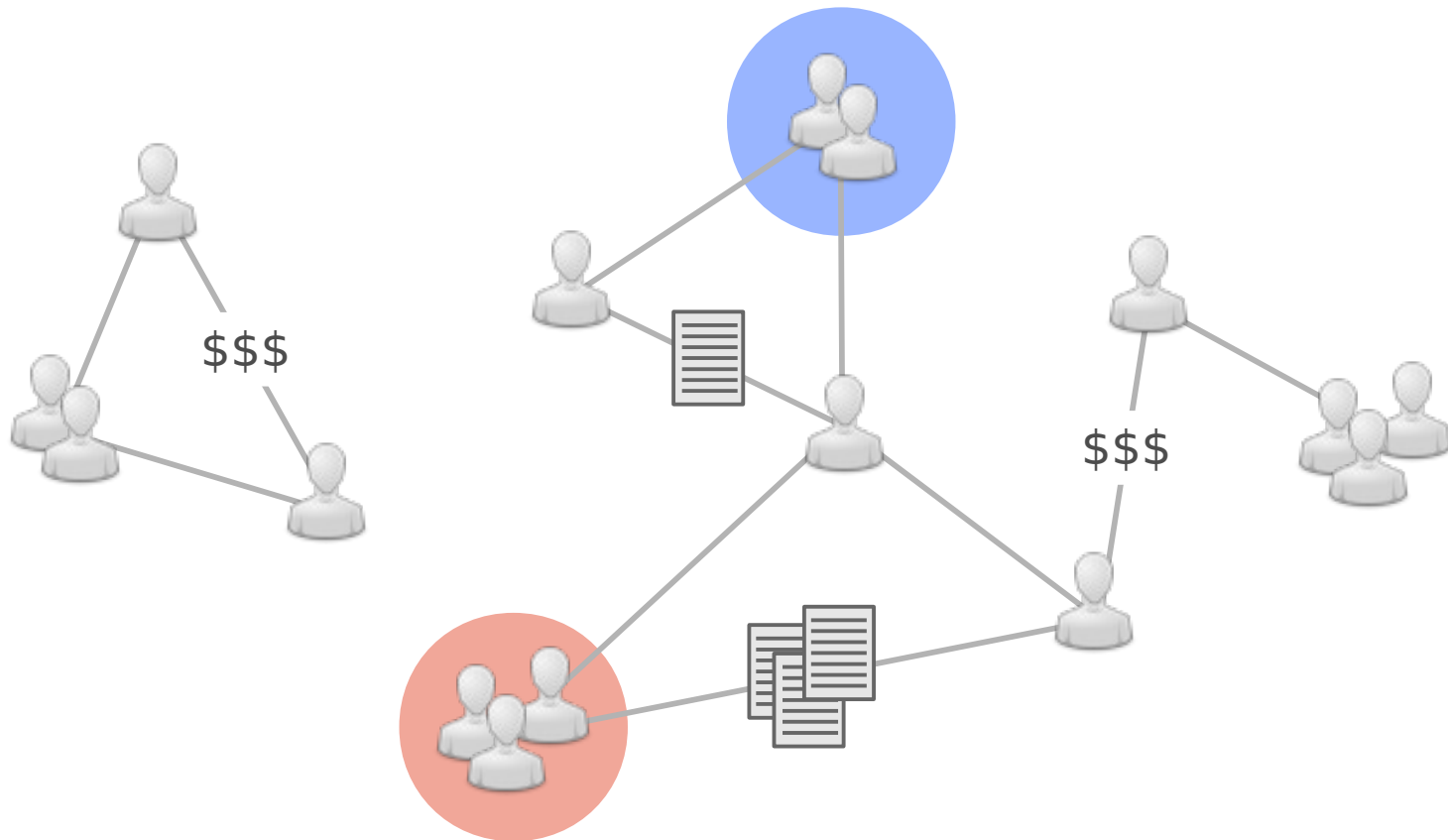
Social Processes: Structure



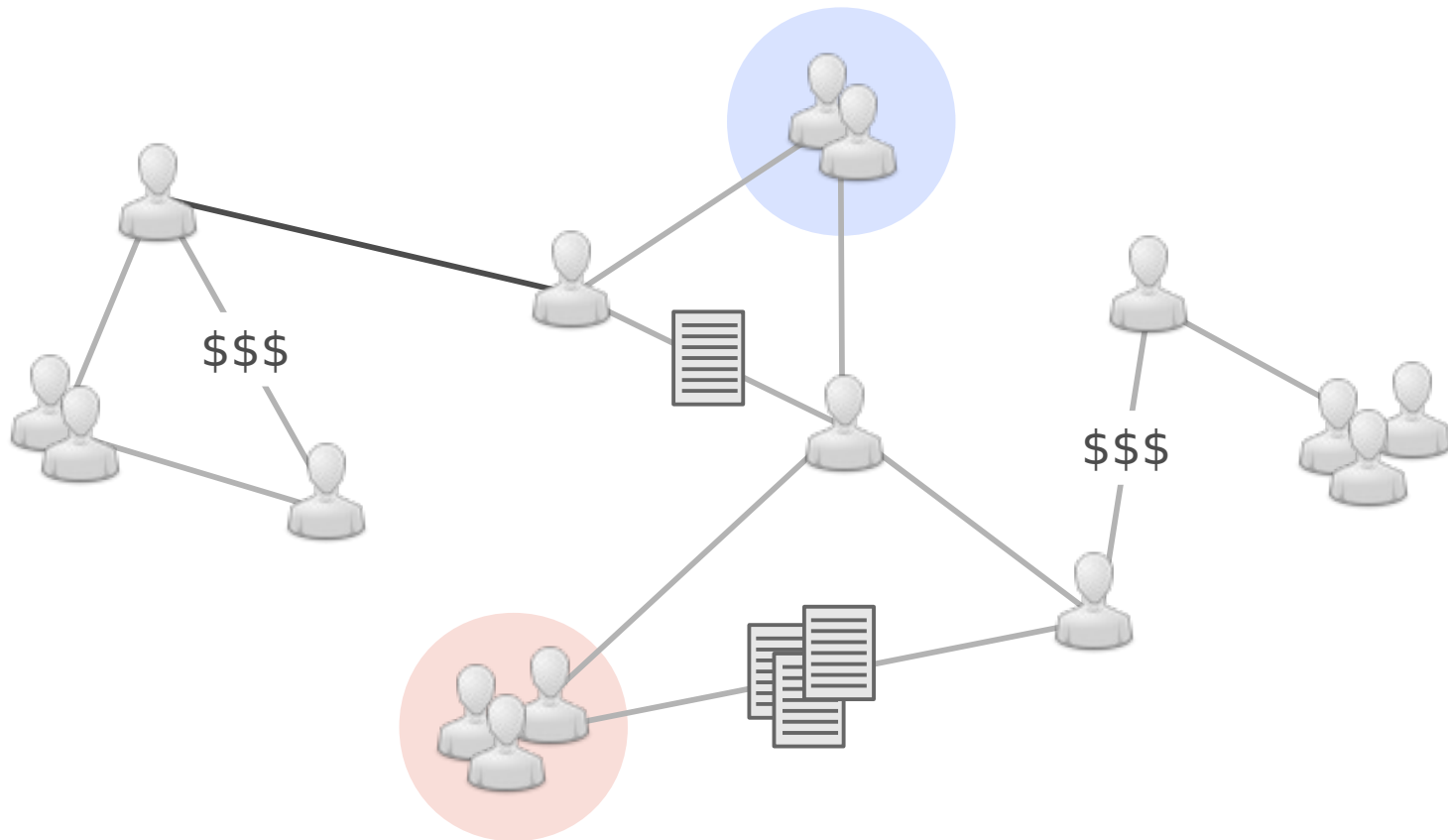
Social Processes: Content



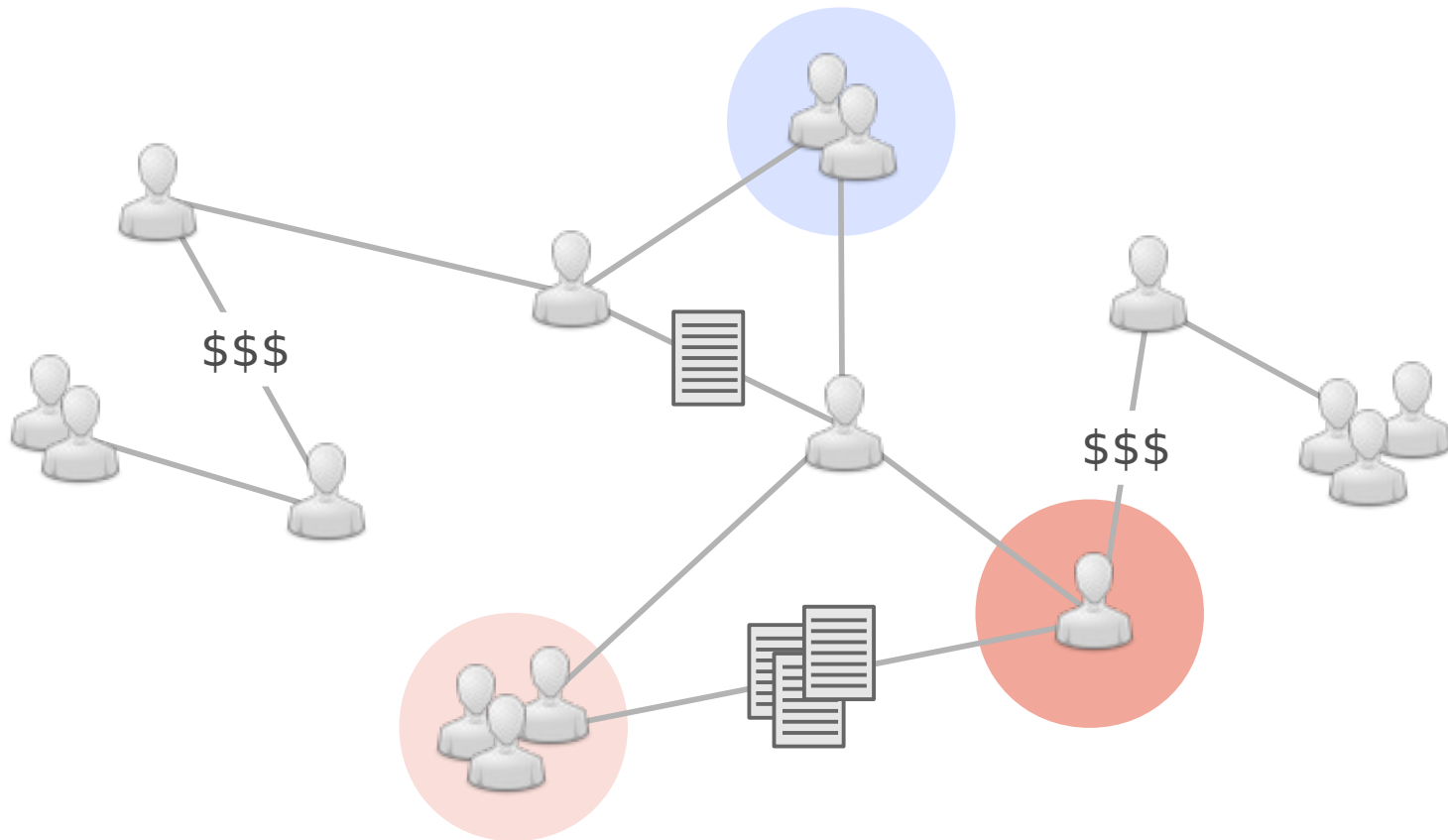
Social Processes: Dynamics



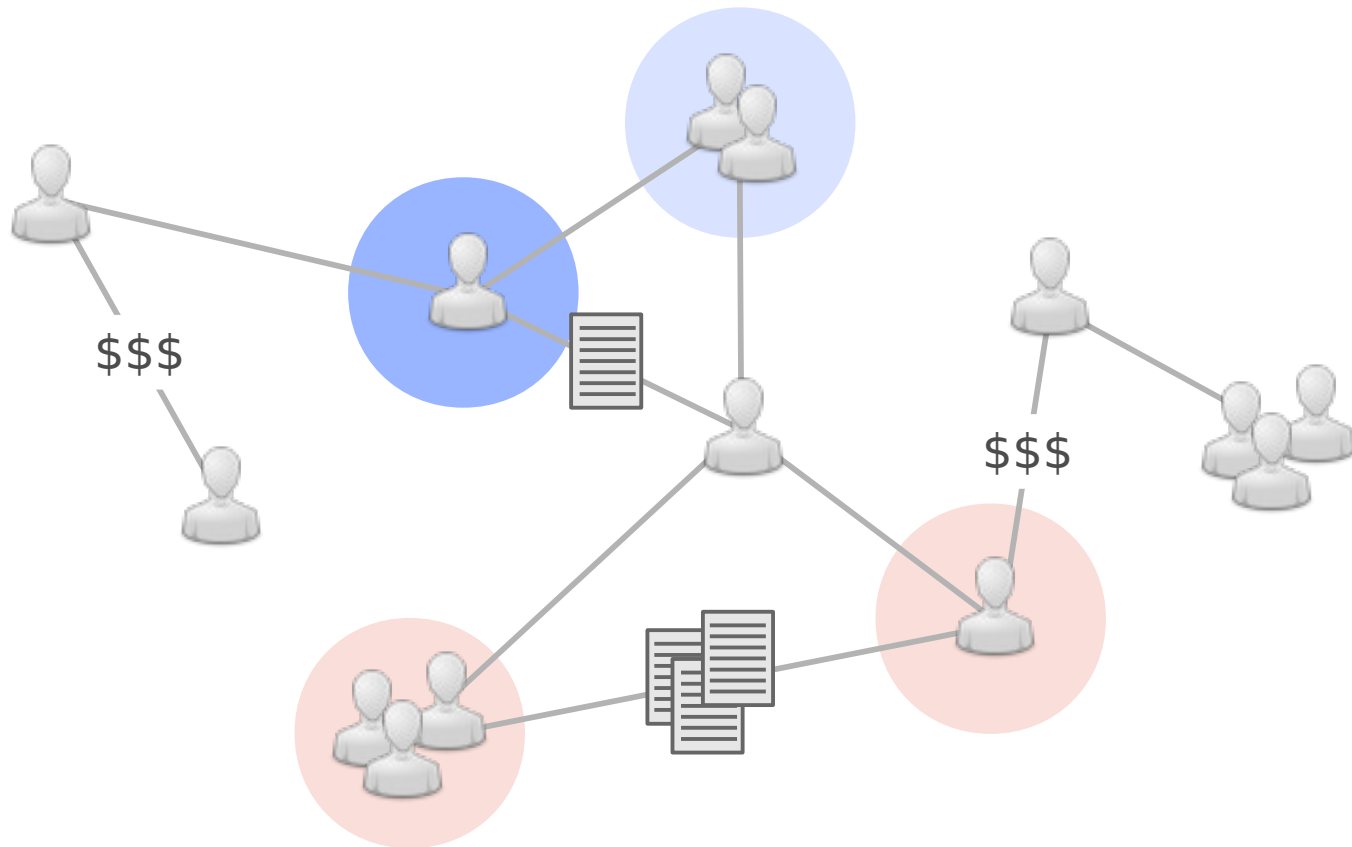
Social Processes: Dynamics



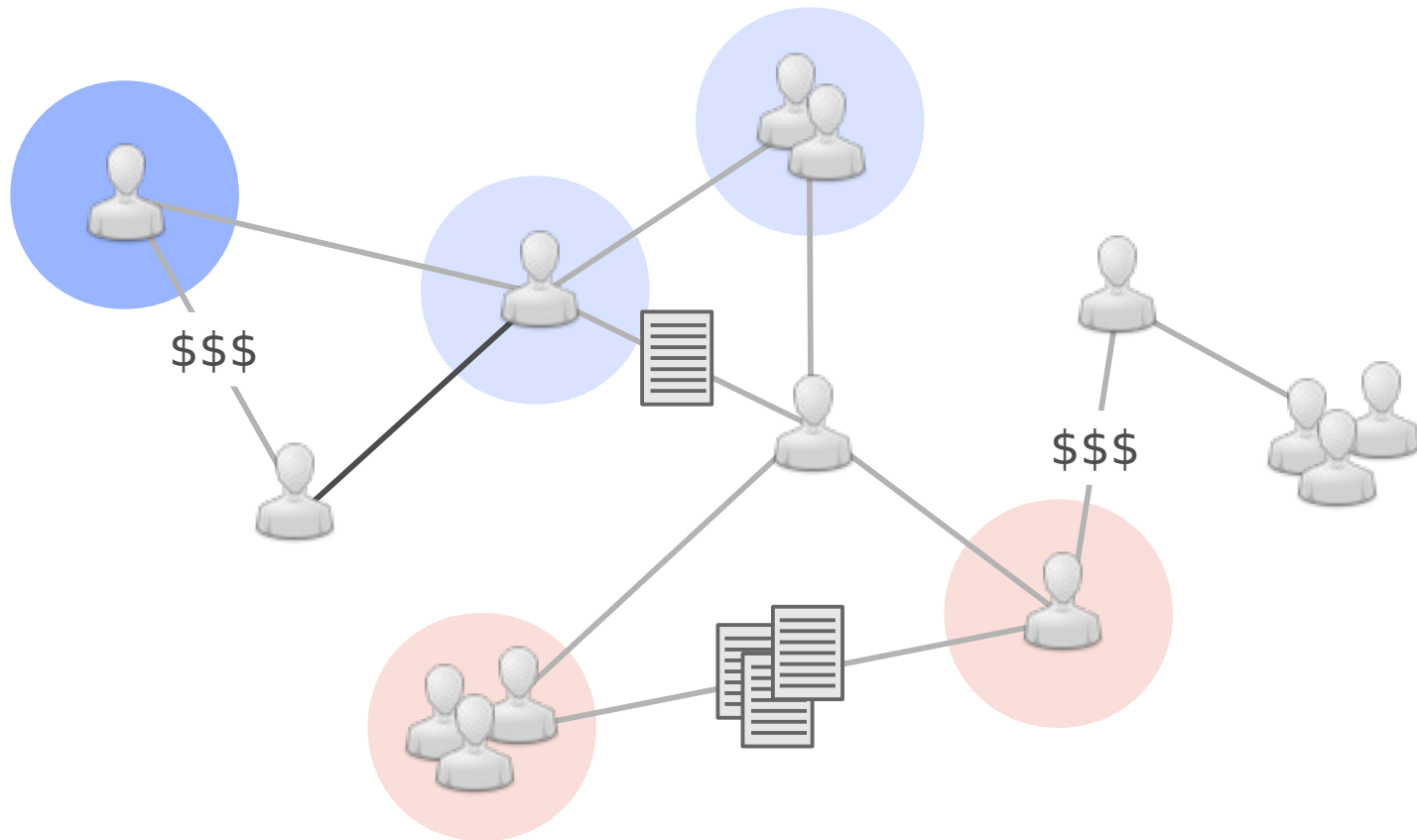
Social Processes: Dynamics



Social Processes: Dynamics



Social Processes: Dynamics



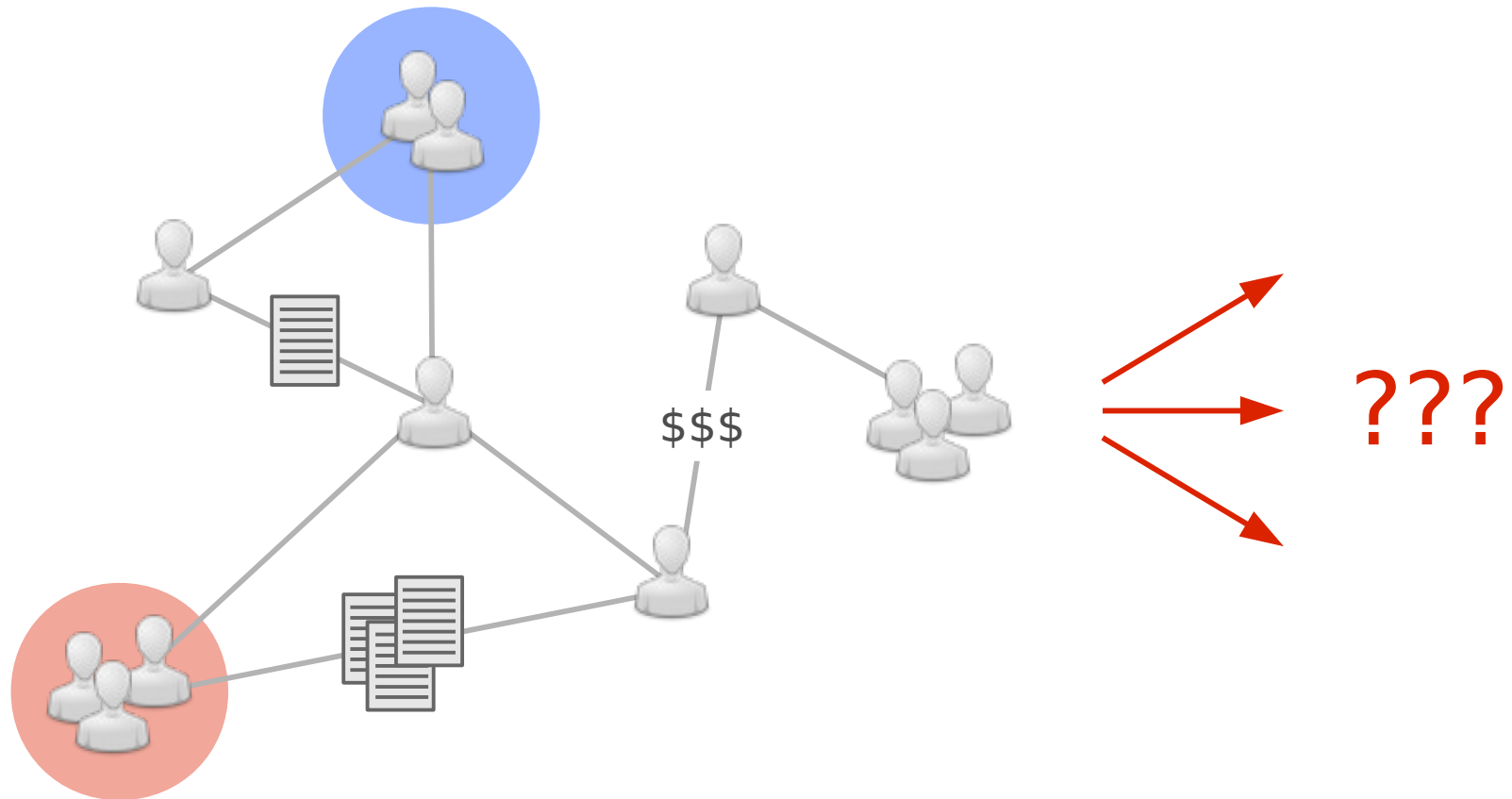
Modeling Social Processes



“Policy-makers or computer scientists may be interested in finding the needle in the haystack (such as a potential terrorist threat or the right web page to display from a search), but social scientists are more commonly interested in characterizing the haystack.”

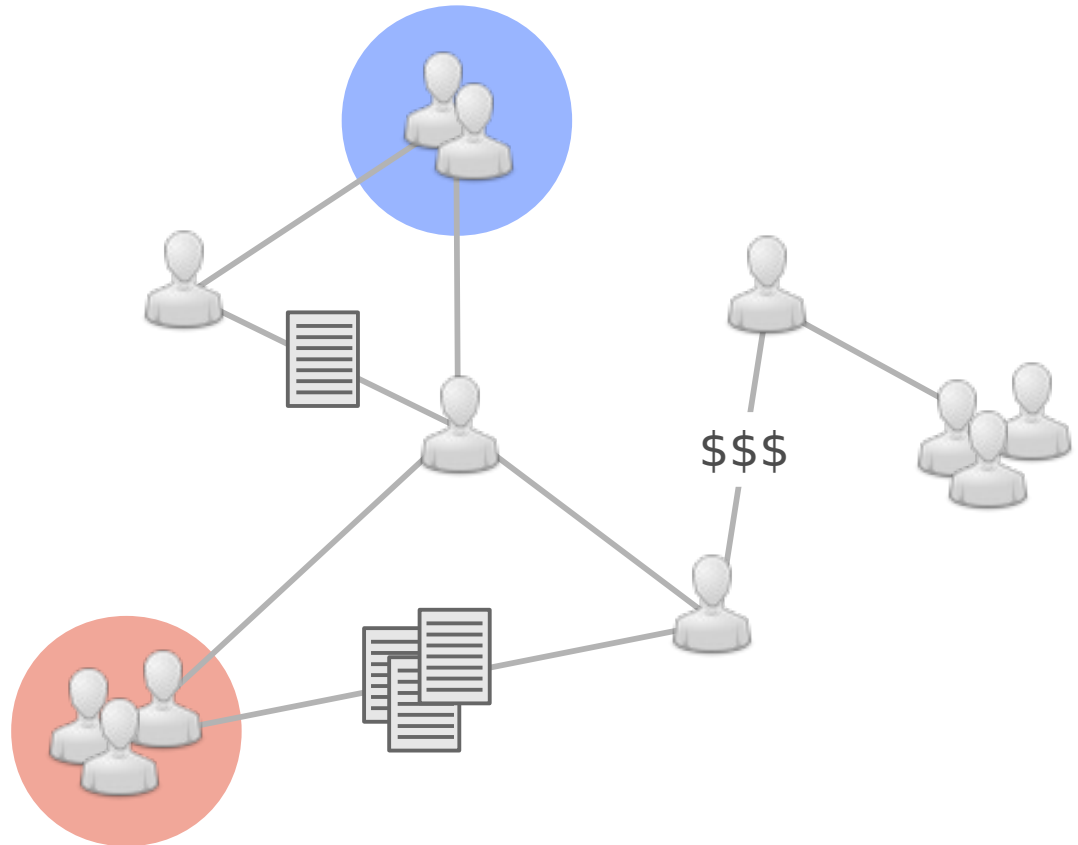
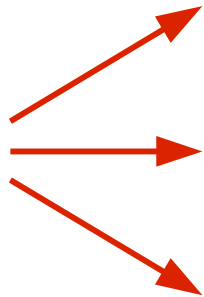
— King & Hopkins, 2010

Predictive Analyses

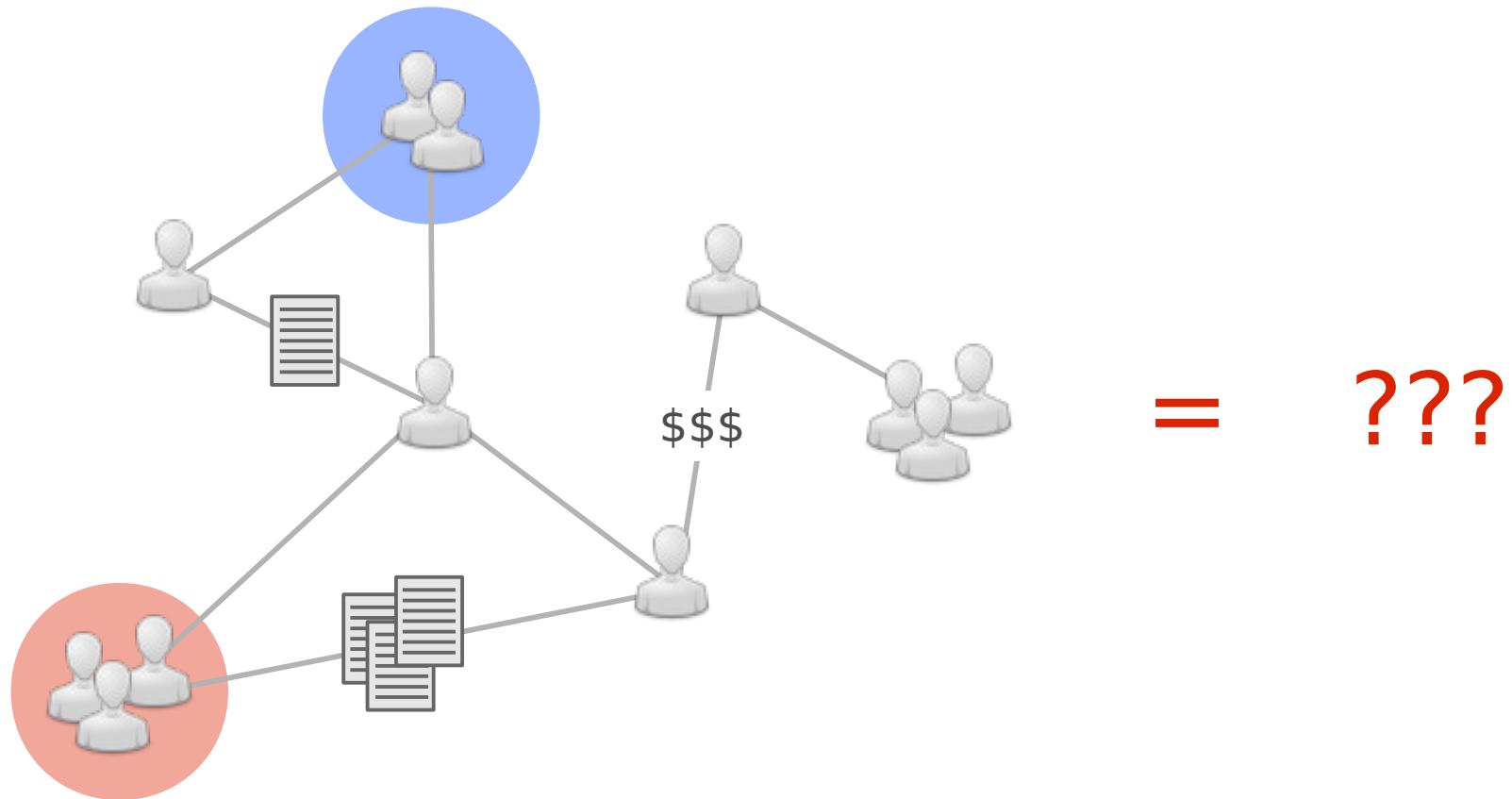


Explanatory Analyses

???



Exploratory Analyses



Bayesian Latent Variable Models

- Modeling challenges:
 - Aggregating and representing large data sets
 - Handling data from sources with disparate emphases
 - Efficiently reasoning under uncertain information
- Bayesian latent (i.e., hidden) variable models:
 - Appropriate for prediction, explanation, and exploration
 - Interpretable structure, not “black-box” models
 - Powerful, flexible, widely applicable...

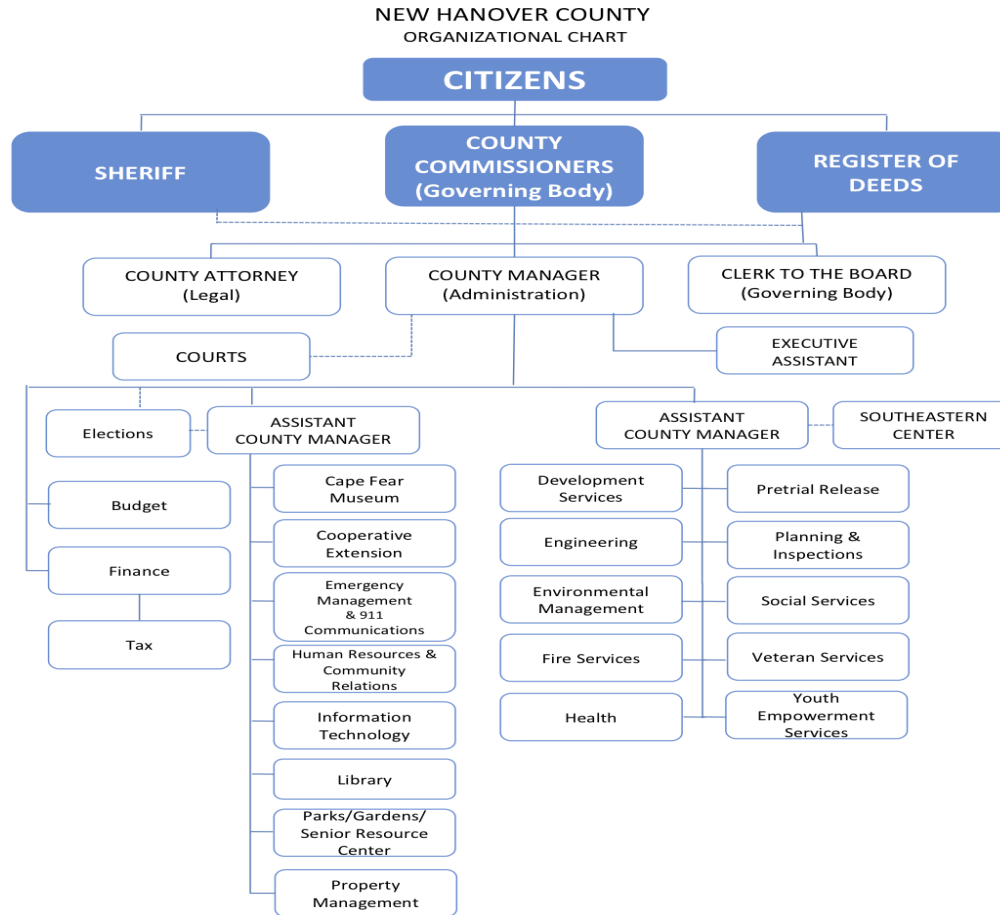
Communication Networks



Communication Networks



Communication Networks



Observing Communication Networks

The screenshot shows a Gmail interface with a search bar containing 'from:adam.kalai@microsoft.com'. Below the search bar are buttons for 'Mail', a checkbox, a refresh icon, and a 'More' dropdown. On the left is a sidebar with a 'COMPOSE' button and a list of folders: Inbox, Starred, Important, Sent Mail, Drafts (2), All Mail, Trash, Admin, Funding, Research, Service, Teaching, Travel, and More. The main area displays a contact card for Adam Kalai with an 'Add to circles' button. Below the card is a list of four email threads, each with a checkbox, a star icon, a right-pointing arrow, the sender 'Adam, Hanna', a count in parentheses, and a subject line. The threads are: 1) 'Travel/2013-05-01 Boston' (3 emails), 2) 'Travel/2013-05-01 Boston' (6 emails), 3) 'Admin/MLDS Research/Email' (13 emails), and 4) 'Travel/2013-05-01 Boston' (5 emails). At the bottom, it shows '17% full Using 4.3 GB of your 25 GB' and copyright information for 2013 Google.

Google from:adam.kalai@microsoft.com

Mail ↕ More ▾

COMPOSE

Adam Kalai ▾
Add to circles

✉ Adam.Kalai@microsoft.com

<input type="checkbox"/>	☆	»	Adam, Hanna (3)	Travel/2013-05-01 Boston	NE
<input type="checkbox"/>	☆	»	Adam, Hanna (6)	Travel/2013-05-01 Boston	RE
<input type="checkbox"/>	☆	»	Adam .. Juston, Peter (13)	Admin/MLDS	Research/Email
<input type="checkbox"/>	☆	»	Adam, Hanna (5)	Travel/2013-05-01 Boston	Sp

17% full
Using 4.3 GB of your 25 GB

©2013 Google - [Terms of Service](#)
[Program Policies](#)

Powered by **GOO**

Structure and Content

Subject: New Hanover County Public Safety Talk Groups
From: "Lee, Warren" <WLee@nhcgov.com>
To: "Pope, Troy W." <twpope@ncshp.org>
Cc: ...

Troy,

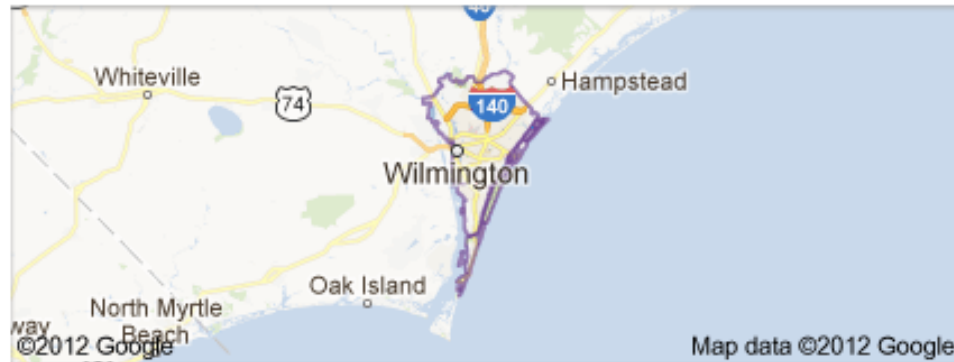
I wanted to give you an update on our progress in moving towards a fully digital public safety radio system in New Hanover County...



New Hanover County, NC

New Hanover County

North Carolina



New Hanover County is one of 100 counties located in the U.S. state of North Carolina. Though second smallest in area, it is one of the most populous as its county seat, Wilmington, is one of the state's largest cities. [Wikipedia](#)

Area: 328 sq miles (849.5 km²)

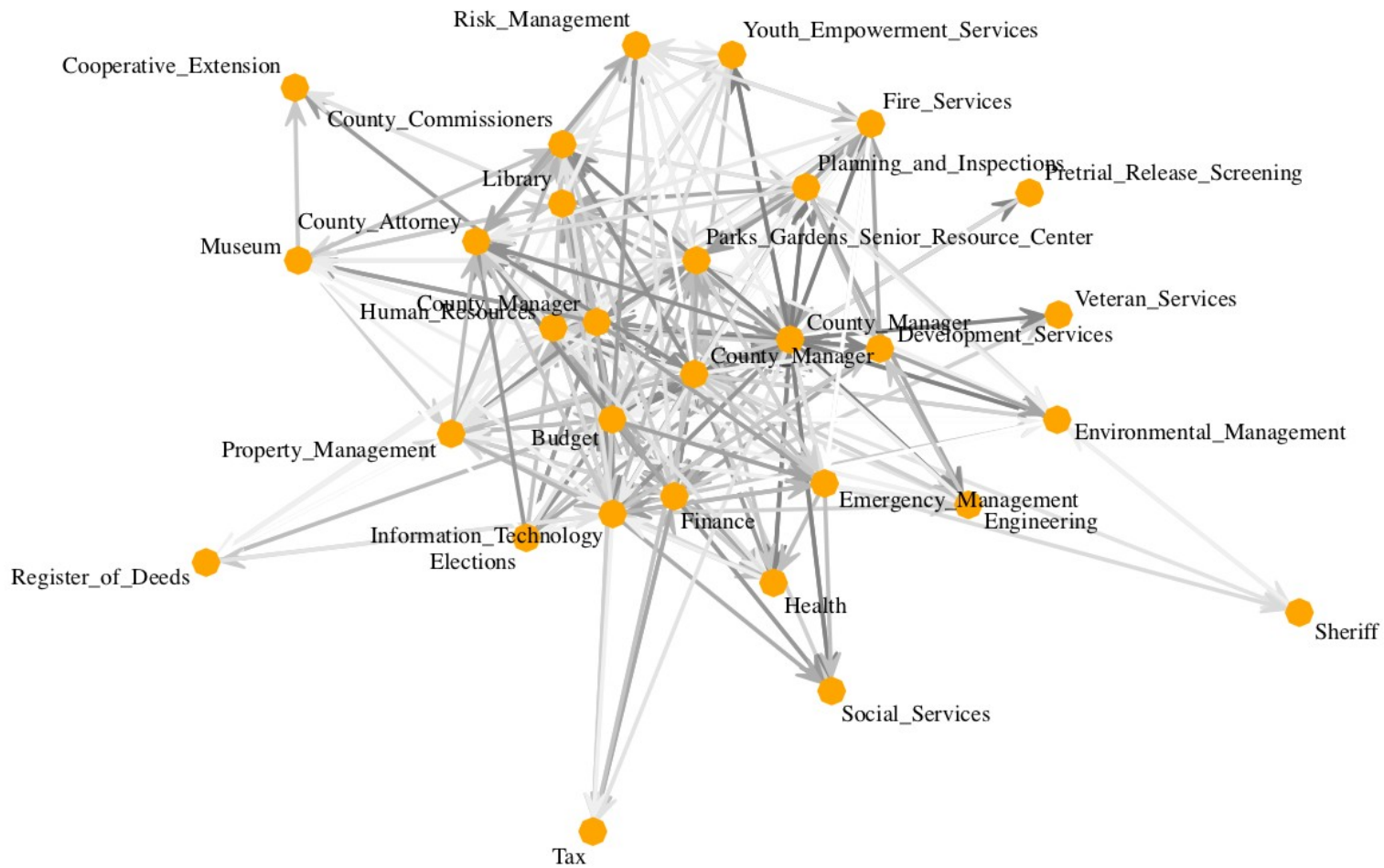
Founded: 1729

Population: 206,189 (2011)

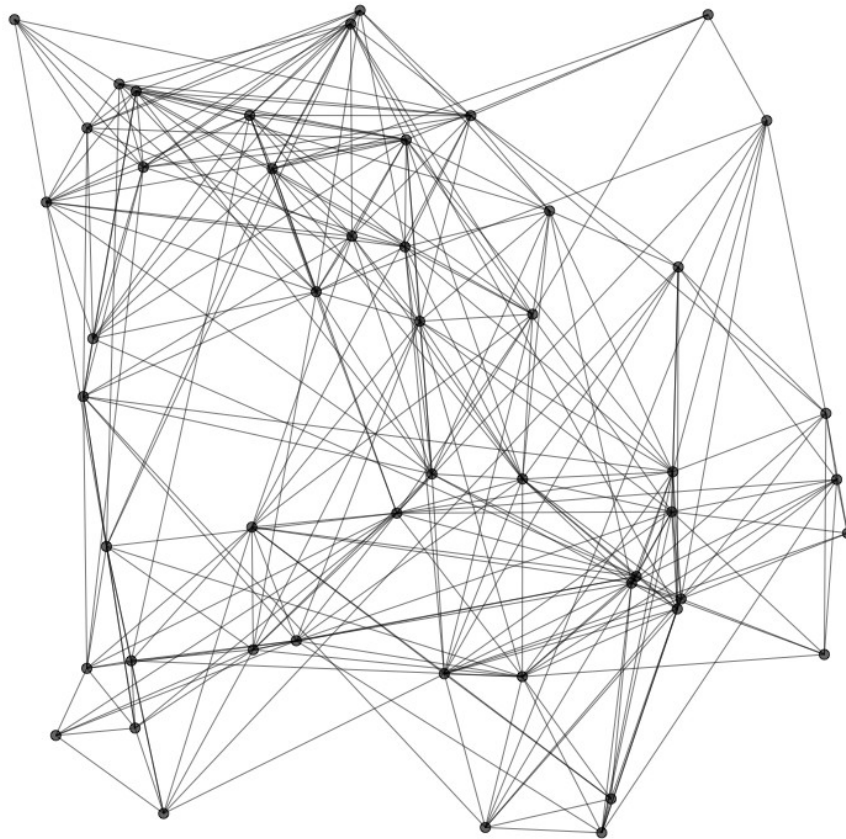
County seat: [Wilmington](#)

[Feedback](#)

NHC Email Network

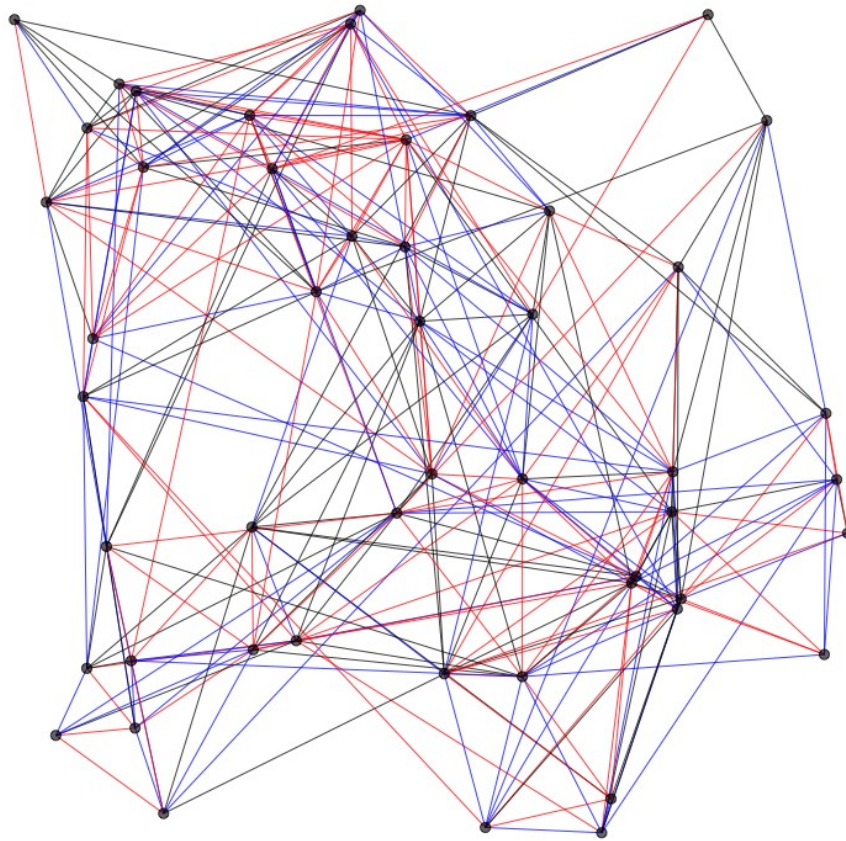


Levels of Granularity



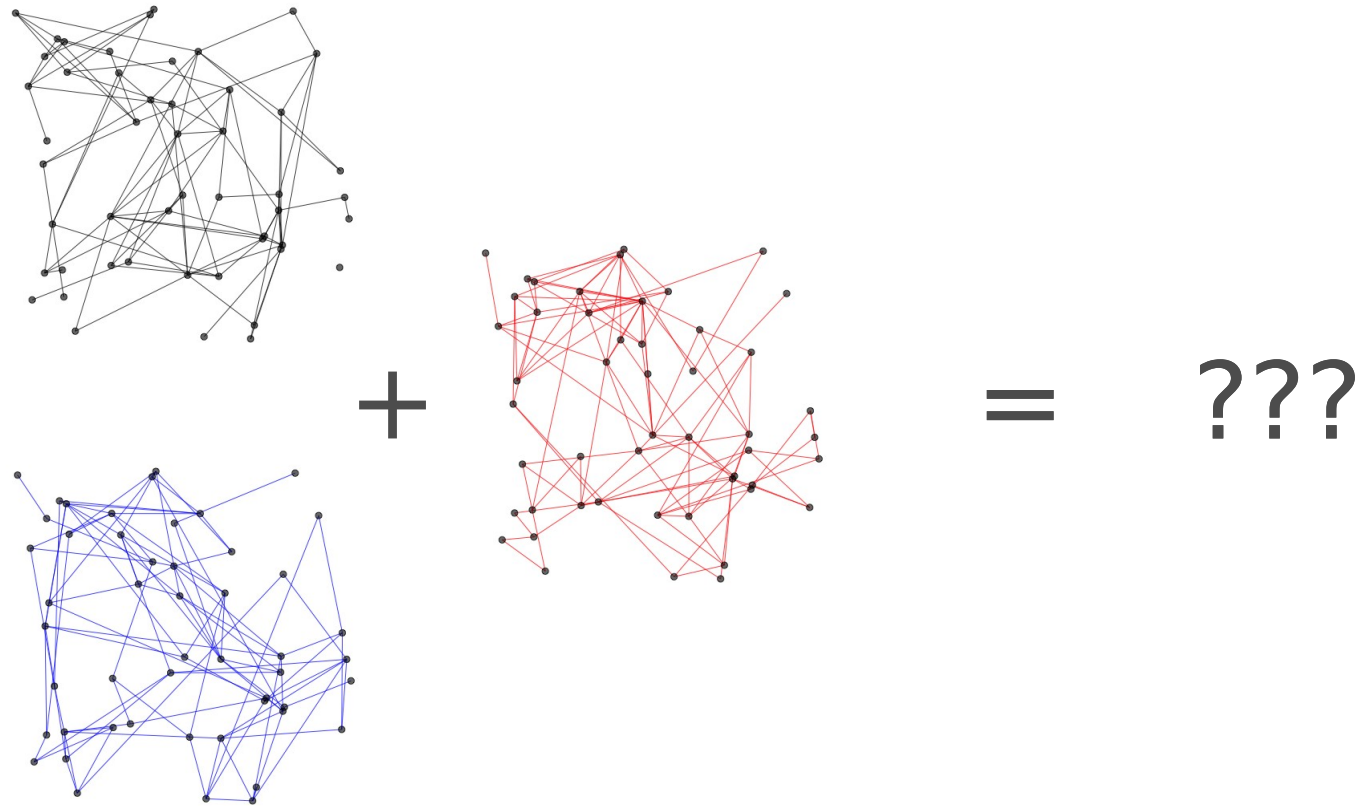
= ???

Levels of Granularity



= ???

Levels of Granularity



Principled Visualization

- Common workflow:
 - Construct a statistical model of observed data
 - Perform post-hoc visualization to draw conclusions about the model and its relationship to the data
- Problem: visualization algorithms can produce visual artifacts that may be misleading
- Solution: visualizations should be directly interpretable in terms of the model and its relationship to the data

Exploring Structure and Content

- Facilitate exploratory analysis of topic-specific communication patterns by learning
 - Topics of communication
 - Topic-specific communication subnetworks
 - Principled visualizations of topic-specific subnetwork
- Draw upon ideas from two well-known frameworks:
 - Statistical topic modeling
 - Latent space network modeling

Topics and Words

gene	ncbi	computer	patent
genome	national	modeling	patenting
dna	information	data	claims
genetic	technology	algorithm	intellectual
genes	database	analyses	property
sequence	molecular	method	rights
human	biology	model	ip
protein	genbank	information	innovation
rna	pubmed	efficient	claim
genomic	references	complexity	claiming
...

Documents and Topics

POLICY FORUM

INTELLECTUAL PROPERTY

Intellectual Property Landscape of the Human Genome

Kyle Jensen and Fiona Murray*

Gene patents are the subject of considerable debate and yet, like the term “gene” itself, the definition of what constitutes a gene patent is fuzzy (1). Nonetheless, gene patents that seem to cause the most controversy are those claiming human protein-encoding nucleotide sequences. This category is the subject of our analysis of the patent landscape of the human genome (2).

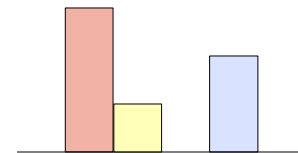
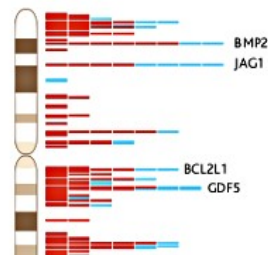
Critics describe the growth in gene sequence patents as an intellectual property (IP) “land grab” over a finite number of human genes (3, 4). They suggest that overly broad patents might block follow-on research (5). Alternatively, gene IP rights may become highly fragmented and cause an anticommens effect, imposing high costs on future innovators and underuse of genomic resources (6). Both situations, critics argue, would increase the costs of genetic diagnostics, slow the development of new medicines, stifle academic research,

tinguishing patents on the human genome from those on other species (23).

Our detailed map was developed using bioinformatics methods to compare nucleotide sequences claimed in U.S. patents to the human genome. Specifically, this map is based on a BLAST (24) homology search linking nucleotide sequences disclosed and claimed in granted U.S. utility patents to the set of protein-encoding messenger RNA transcripts contained in the National Center for Biotechnology Information (NCBI) RefSeq (25) and Gene (26) databases. This method allows us to map gene-oriented IP rights to specific physical loci on the human genome (27) (see figure, right). Our approach is highly specific in its identification of patents that actually claim human nucleotide sequences. However, by limiting the search to patents using the canoni-

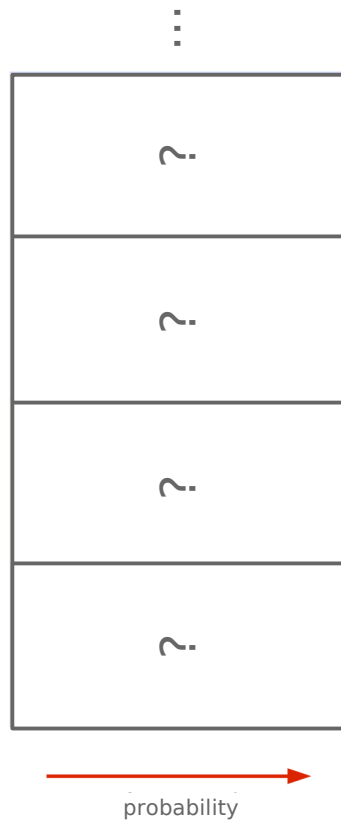
California, Isis Pharmaceuticals, the former SmithKline Beecham, and Human Genome Sciences. The top patent assignee is Incyte Pharmaceuticals/Incyte Genomics, whose IP rights cover 2000 human genes, mainly for use as probes on DNA microarrays.

Although large expanses of the genome are unpatented, some genes have up to 20 patents asserting rights to various gene uses and manifestations including diagnostic uses, single nucleotide polymorphisms (SNPs), cell lines, and constructs containing the gene. The distribution of gene patents was nonuniform (see figure, page 240, top right): Specific regions of the genome are “hot spots” of heavy patent activity, usually with a one-gene-many-patents scenario (see figure, below). Although less common, there were cases in which a single patent claims many genes, typically as complementary DNA probes used on a microarray (see figure, p. 240, bottom).



Latent Dirichlet Allocation

[Blei, Ng & Jordan, '03]



POLICY FORUM

INTELLECTUAL PROPERTY

Intellectual Property Landscape of the Human Genome

Kyle Jensen and Fiona Murray*

Gene patents are the subject of considerable debate and yet, like the term "gene" itself, the definition of what constitutes a gene patent is fuzzy (1). Nonetheless, gene patents that seem to cause the most controversy are those claiming human protein-encoding nucleotide sequences. This category is the subject of our analysis of the patent landscape of the human genome (2). Critics describe the growth in gene sequence patents as an intellectual property (IP) "land grab" over a finite number of human genes (3, 4). They suggest that overly broad patents might block follow-on research (5). Alternatively, gene IP rights may become highly fragmented and cause an anticommons effect, imposing high costs on future innovators and underuse of genomic resources (6). Both situations, critics argue, would increase the costs of genetic diagnostics, slow the development of new medicines, stifle academic research, distinguishing patents on the human genome from those on other species (23).

Our detailed map was developed using bioinformatics methods to compare nucleotide sequences claimed in U.S. patents to the human genome. Specifically, this map is based on a BLAST (24) homology search linking nucleotide sequences disclosed and claimed in granted U.S. utility patents to the set of protein-encoding messenger RNA transcripts contained in the National Center for Biotechnology Information (NCBI) RefSeq (25) and Gene (26) databases. This method allows us to map gene-oriented IP rights to specific physical loci on the human genome (27) (see figure, right). Our approach is highly specific in its identification of patents that actually claim human nucleotide sequences. However, by limiting the search to patents using the canoni-

California, Isis Pharmaceuticals, the former SmithKline Beecham, and Human Genome Sciences. The top patent assignee is Incyte Pharmaceuticals/Incyte Genomics, whose IP rights cover 2000 human genes, mainly for use as probes on DNA microarrays.

Although large expanses of the genome are unpatented, some genes have up to 20 patents asserting rights to various gene uses and manifestations including diagnostic uses, single nucleotide polymorphisms (SNPs), cell lines, and constructs containing the gene. The distribution of gene patents was nonuniform (see figure, page 240, top right): Specific regions of the genome are "hot spots" of heavy patent activity, usually with a one-gene-many-patents scenario (see figure, below). Although less common, there were cases in which a single patent claims many genes, typically as complementary DNA probes used on a microarray (see figure, p. 240, bottom).

?

Individuals and Latent Spaces

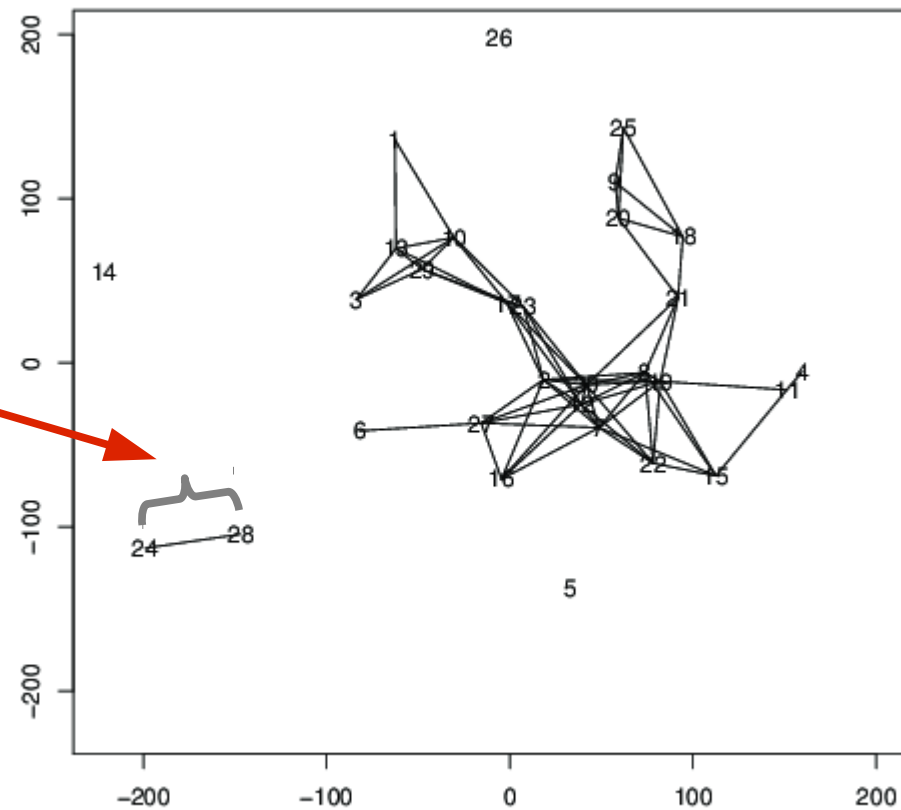
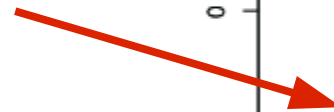
every individual
is associated
with a position
in latent space



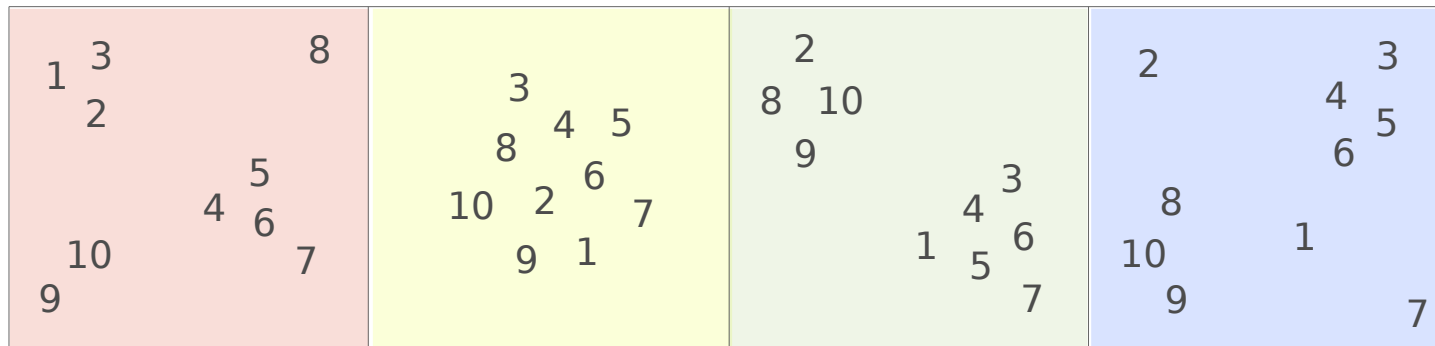
Latent Space Network Model

[Hoff et al., '02]

probability of communication depends on distance in latent space



Topics and Spaces



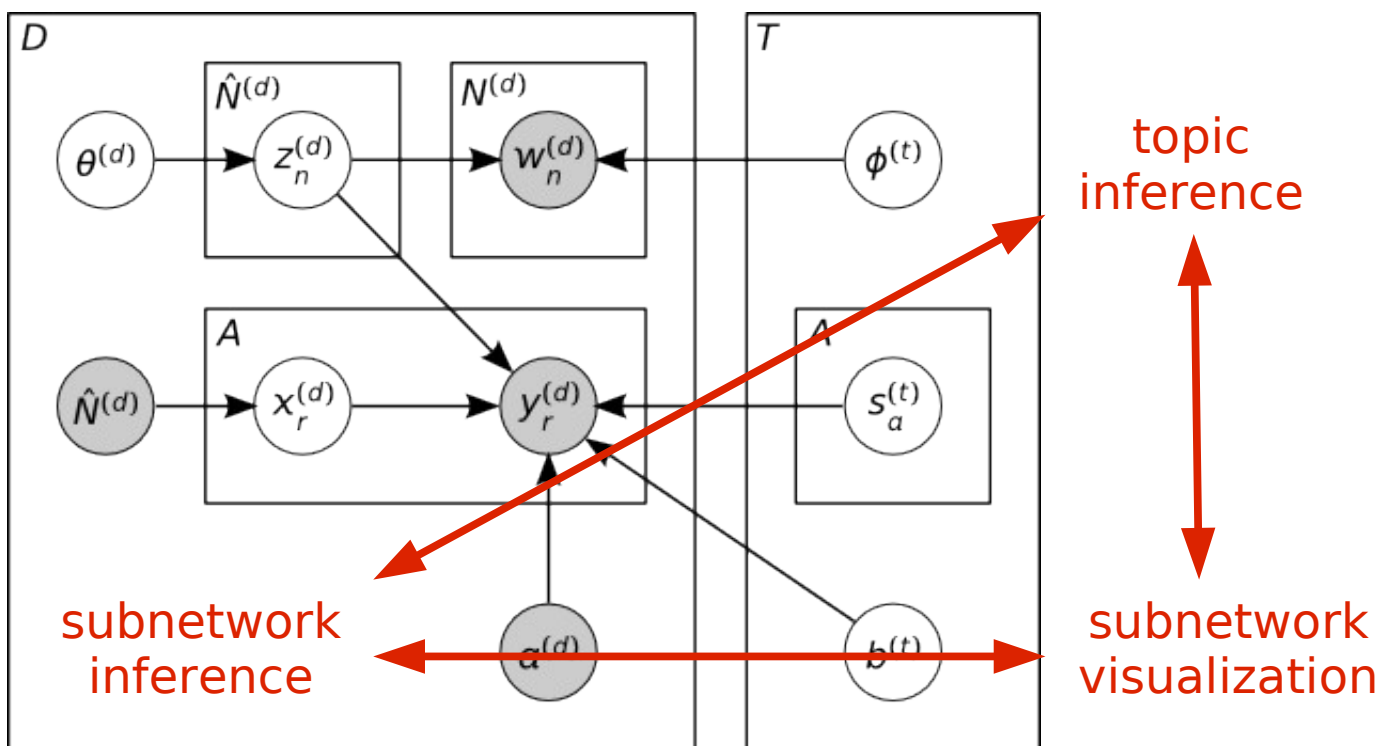
gene genome dna genetic ...	ncbi national information technology ...	computer modeling data algorithm ...	patent patenting claims intellectual ...
---	--	--	--

A New Model...

[Krafft et al., '12]

- Model email content using LDA
- Model recipients using topic-specific latent spaces
- Generative process:
 - Generate topics and topic-specific latent spaces
 - Generate document-specific topic distributions
 - Generate recipients using latent spaces
 - Generate words using topics

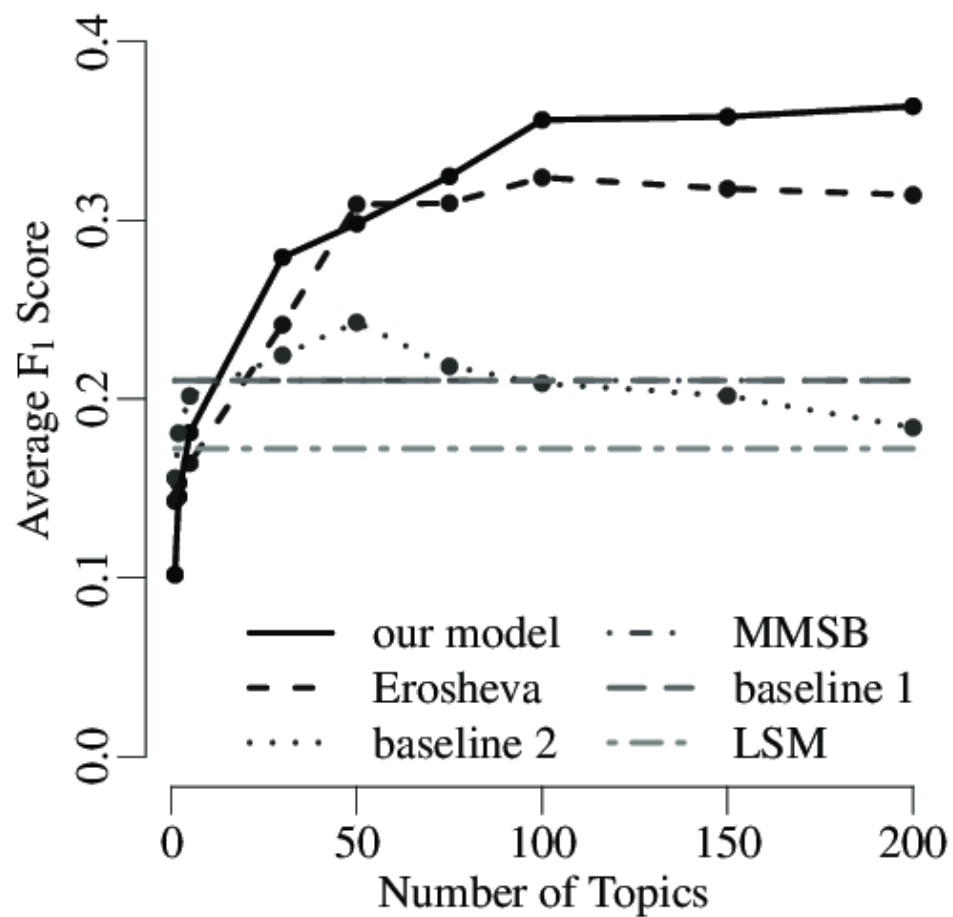
Graphical Model



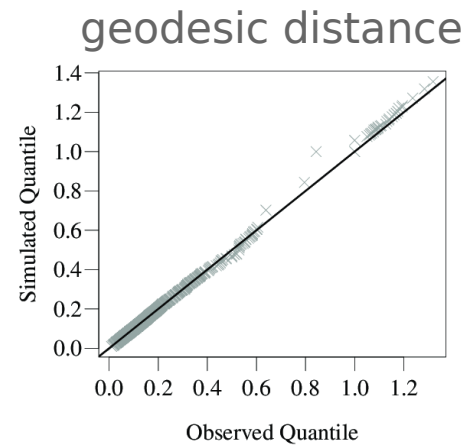
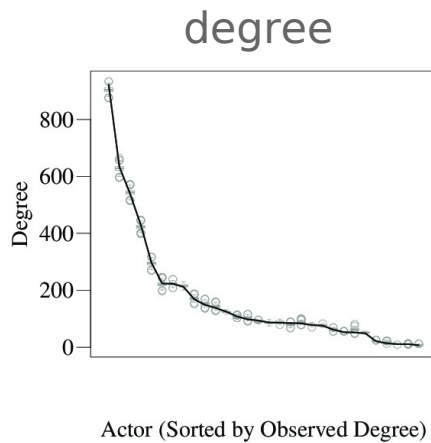
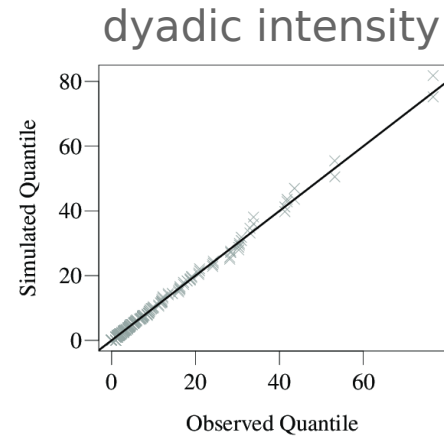
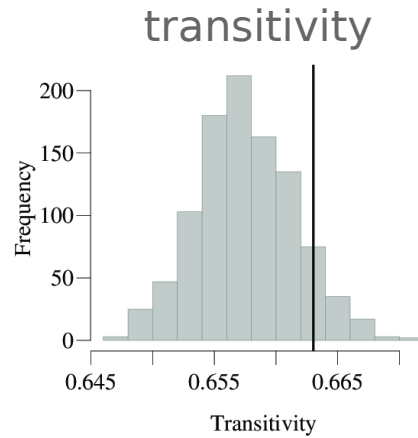
Experimental Evaluation

- Quantitative model validation:
 - Link prediction performance vs. baselines
 - Posterior predictive checks
 - Topic coherence vs. LDA
- Exploratory analysis:
 - Modularity: disconnected components
 - Assortativity: components of a single “type”

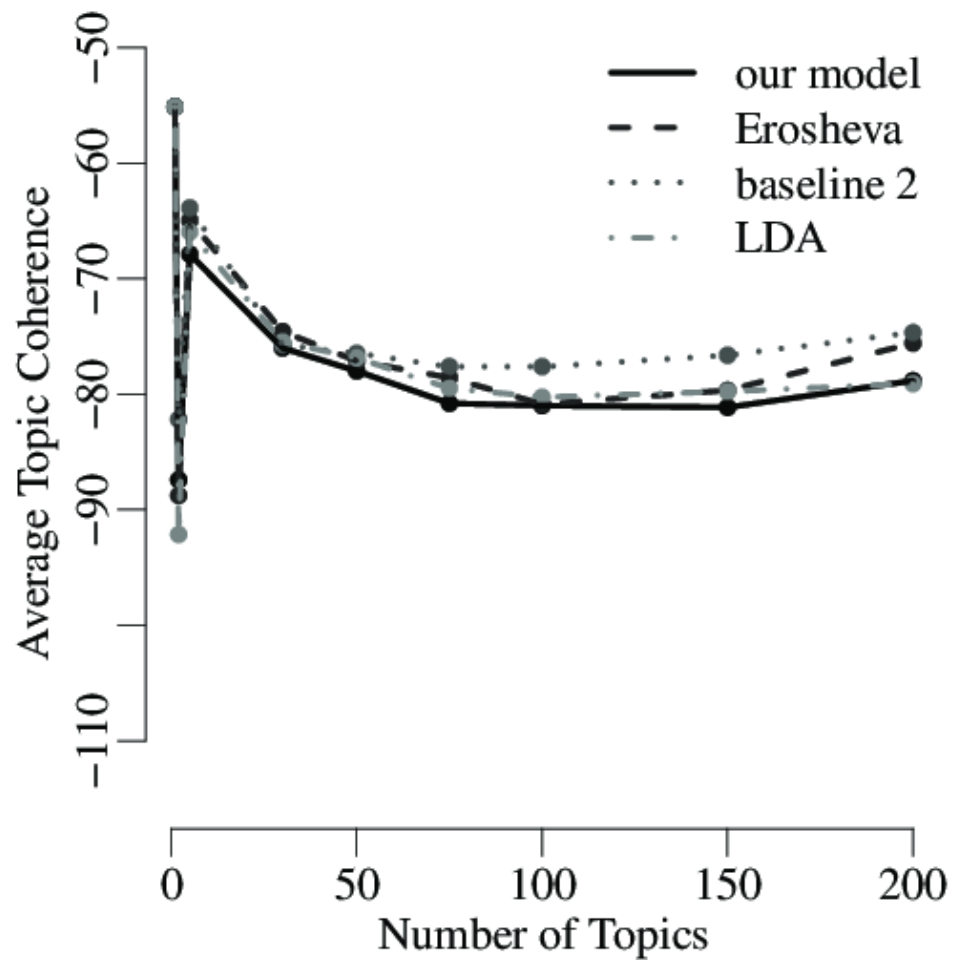
Link Prediction



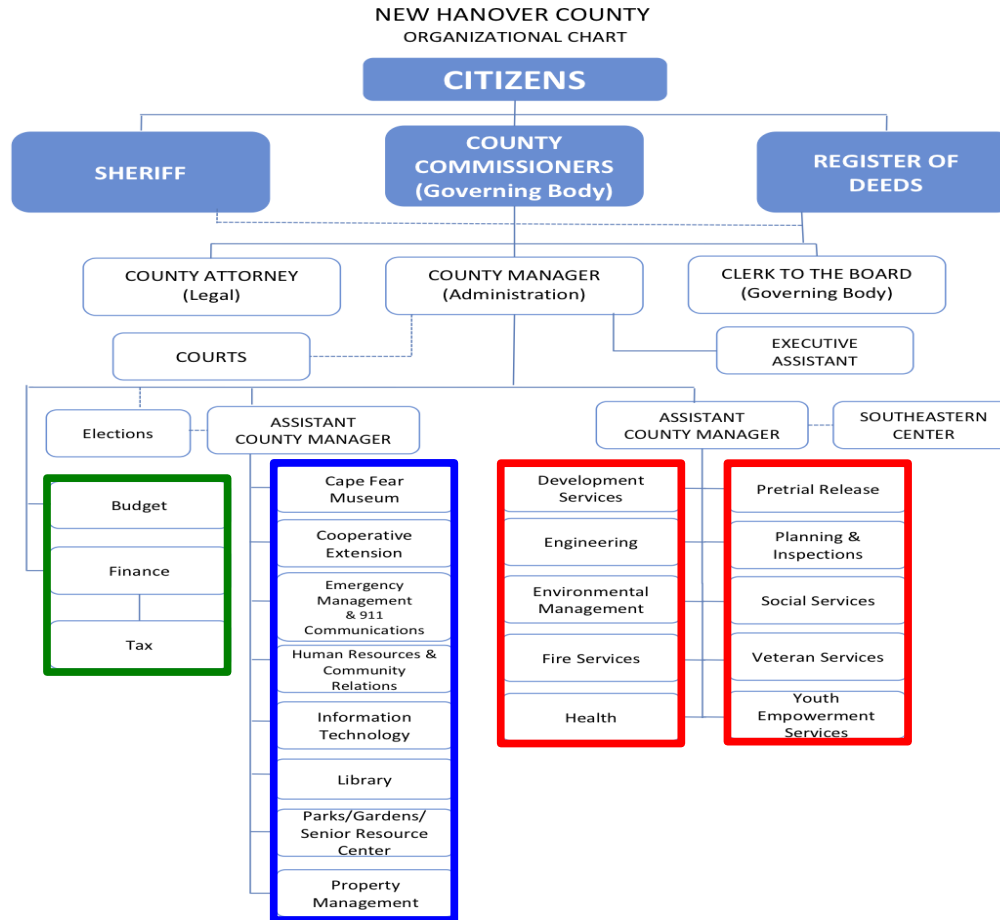
Posterior Predictive Checks



Topic Coherence



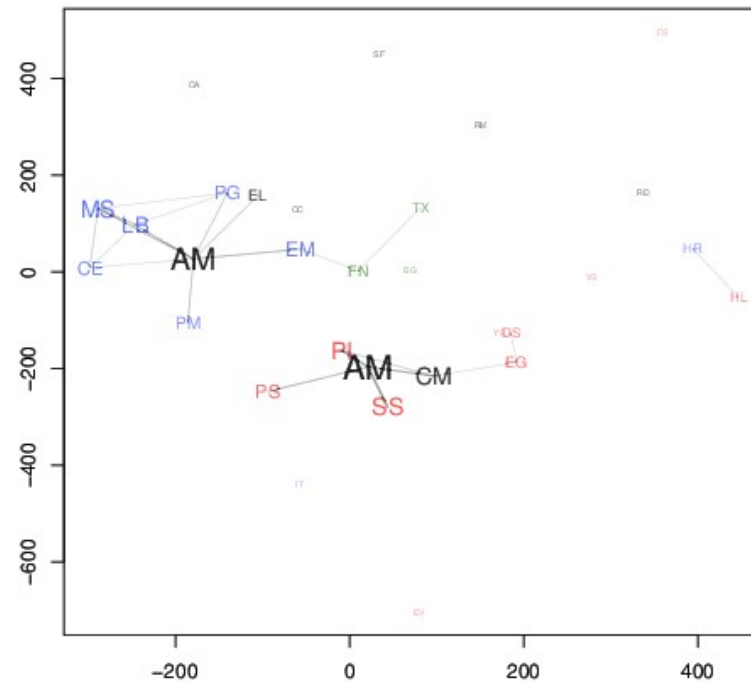
Organization Structure



High Modularity, High Assortativity

Assistant County Manager	AM
Budget	BG
Cooperative Extension	CE
County Attorney	CA
County Commissioners	CC
County Manager	CM
Development Services	DS
Elections	EL
Emergency Management	EM
Engineering	EG
Environmental Management	EV
Finance	FN
Fire Services	FS
Health	HL
Human Resources	HR
Information Technology	IT
Library	LB
Museum	MS
Parks and Gardens	PG
Planning and Inspections	PI
Pretrial Release Screening	PS
Property Management	PM
Register of Deeds	RD
Risk Management	RM
Sheriff	SF
Social Services	SS
Tax	TX
Veteran Services	VS
Youth Empowerment Services	YS

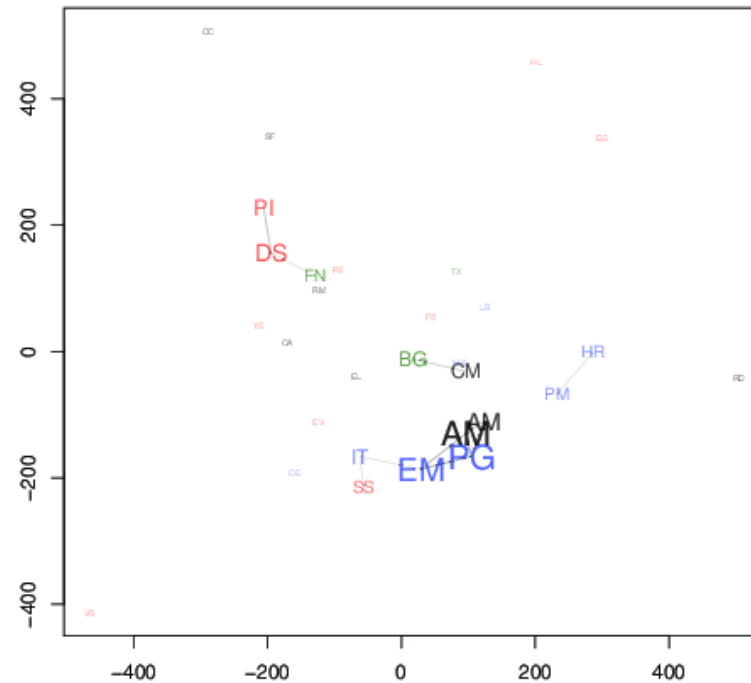
Meeting Scheduling
meeting march board agenda week



High Modularity, Low Assortativity

Assistant County Manager	AM
Budget	BG
Cooperative Extension	CE
County Attorney	CA
County Commissioners	CC
County Manager	CM
Development Services	DS
Elections	EL
Emergency Management	EM
Engineering	EG
Environmental Management	EV
Finance	FN
Fire Services	FS
Health	HL
Human Resources	HR
Information Technology	IT
Library	LB
Museum	MS
Parks and Gardens	PG
Planning and Inspections	PI
Pretrial Release Screening	PS
Property Management	PM
Register of Deeds	RD
Risk Management	RM
Sheriff	SF
Social Services	SS
Tax	TX
Veteran Services	VS
Youth Empowerment Services	YS

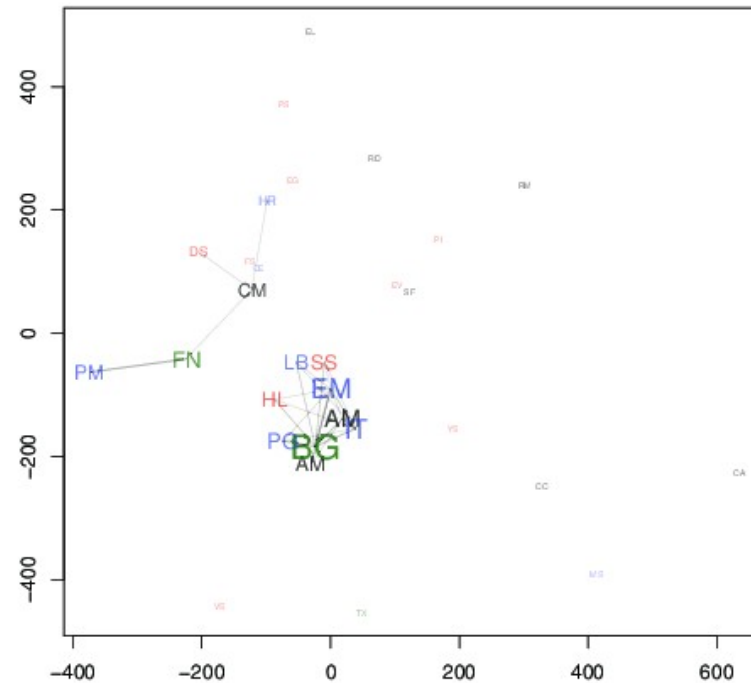
Public Signage
change signs sign process ordinance



Low Modularity, Low Assortativity

Assistant County Manager	AM
Budget	BG
Cooperative Extension	CE
County Attorney	CA
County Commissioners	CC
County Manager	CM
Development Services	DS
Elections	EL
Emergency Management	EM
Engineering	EG
Environmental Management	EV
Finance	FN
Fire Services	FS
Health	HL
Human Resources	HR
Information Technology	IT
Library	LB
Museum	MS
Parks and Gardens	PG
Planning and Inspections	PI
Pretrial Release Screening	PS
Property Management	PM
Register of Deeds	RD
Risk Management	RM
Sheriff	SF
Social Services	SS
Tax	TX
Veteran Services	VS
Youth Empowerment Services	YS

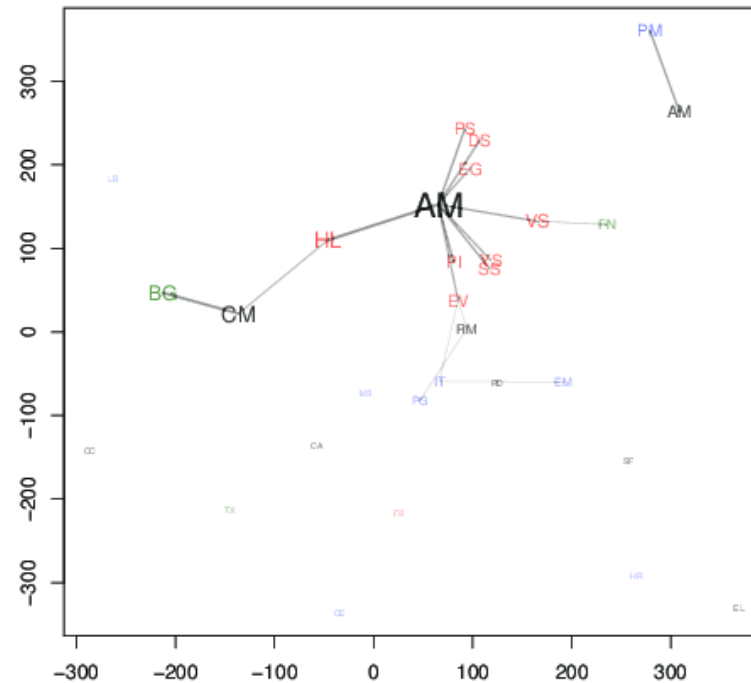
Public Relations
city breakdown information give



Low Modularity, High Assortativity

Assistant County Manager	AM
Budget	BG
Cooperative Extension	CE
County Attorney	CA
County Commissioners	CC
County Manager	CM
Development Services	DS
Elections	EL
Emergency Management	EM
Engineering	EG
Environmental Management	EV
Finance	FN
Fire Services	FS
Health	HL
Human Resources	HR
Information Technology	IT
Library	LB
Museum	MS
Parks and Gardens	PG
Planning and Inspections	PI
Pretrial Release Screening	PS
Property Management	PM
Register of Deeds	RD
Risk Management	RM
Sheriff	SF
Social Services	SS
Tax	TX
Veteran Services	VS
Youth Empowerment Services	YS

Broadcast Messages
fw fyi bulletin summary week



Take Away Message

- Explanatory and exploratory analyses matter
- Communication networks are important:
 - Critical to all kinds of collaborative problem solving
 - ... but can be hard to directly observe
- Topic-partitioned multinet network embedding:
 - Good model of structure and content
 - Emphasizes principled visualization

Thanks!

Acknowledgements: P. Krafft, J. Moore, B. Desmarais, J. ben-Aaron

wallach@cs.umass.edu
<http://www.cs.umass.edu/~wallach/>