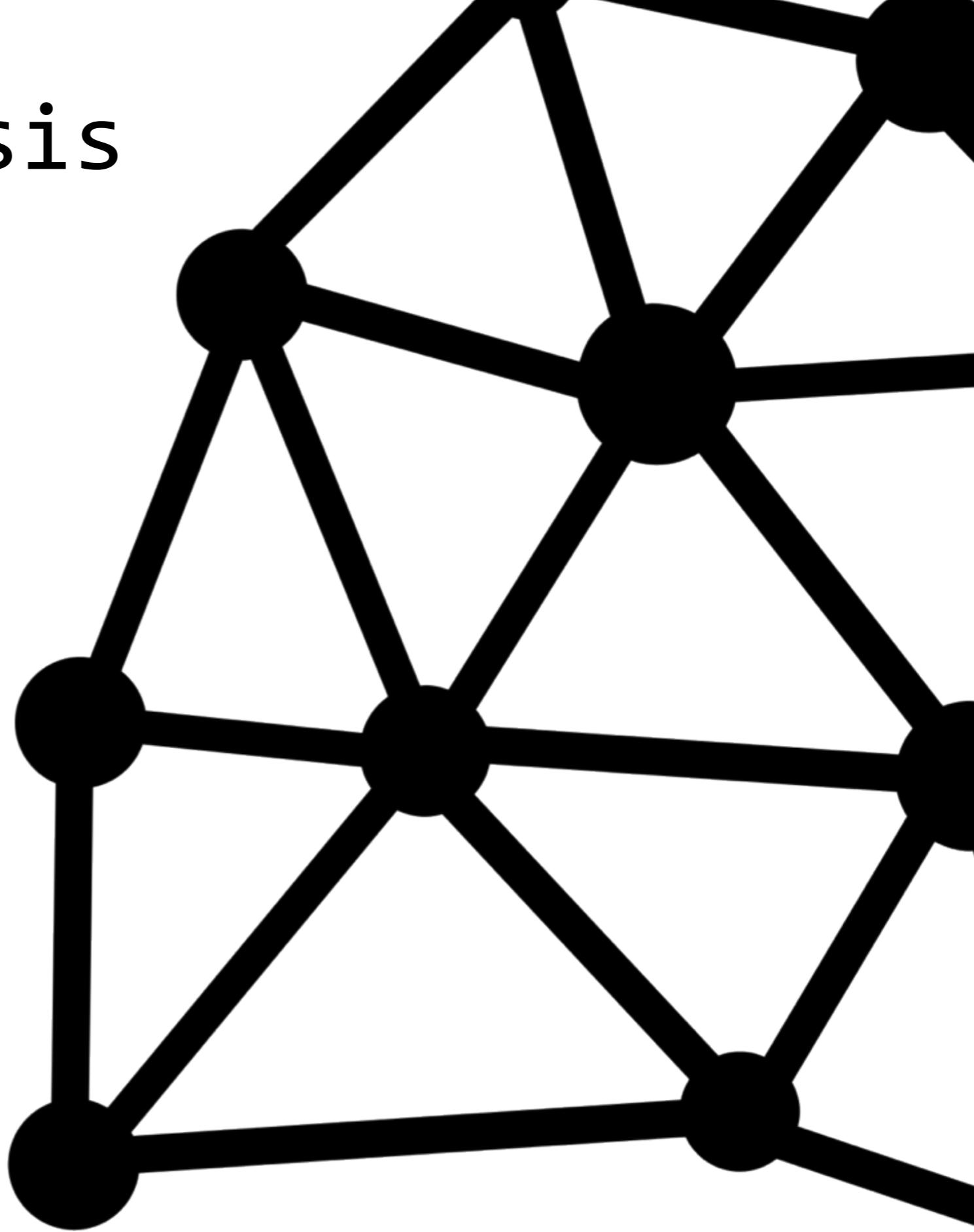
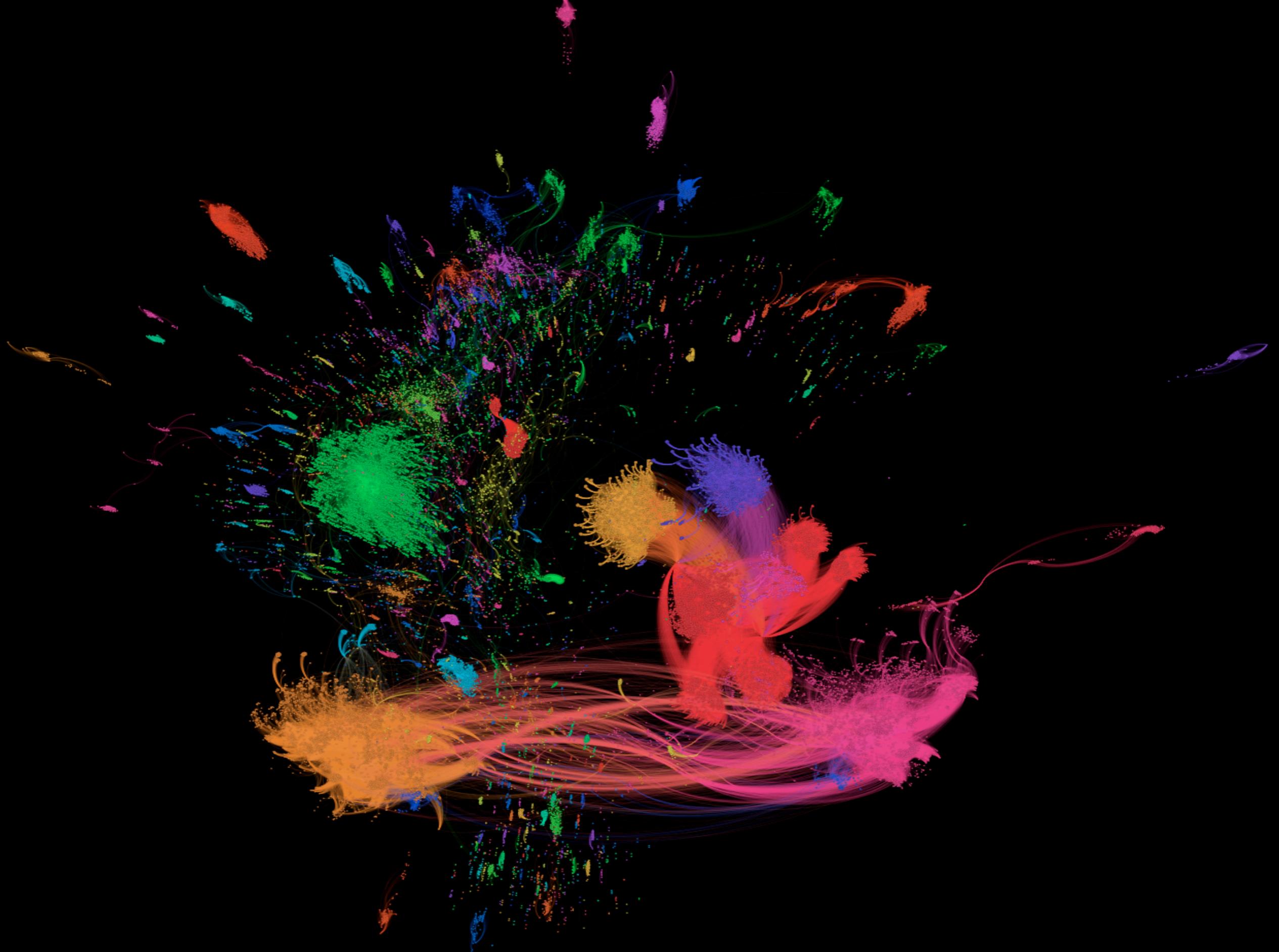


# Network Analysis

## An Introduction

Thomas G. Padilla  
Digital Scholarship Librarian  
@thomasgpadilla





Ian Milligan, Using WarcBase to Generate a Linkgraph of the Wide Web Scrape  
<http://ianmilligan.ca/2015/03/03/using-warcbase-to-generate-a-link-graph-of-the-wide-web-scrape/>



Definitions

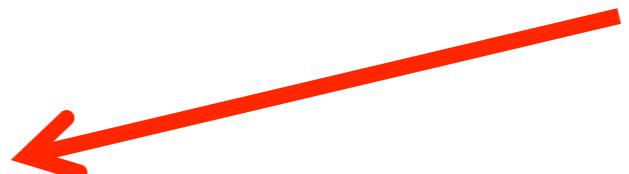
Questions

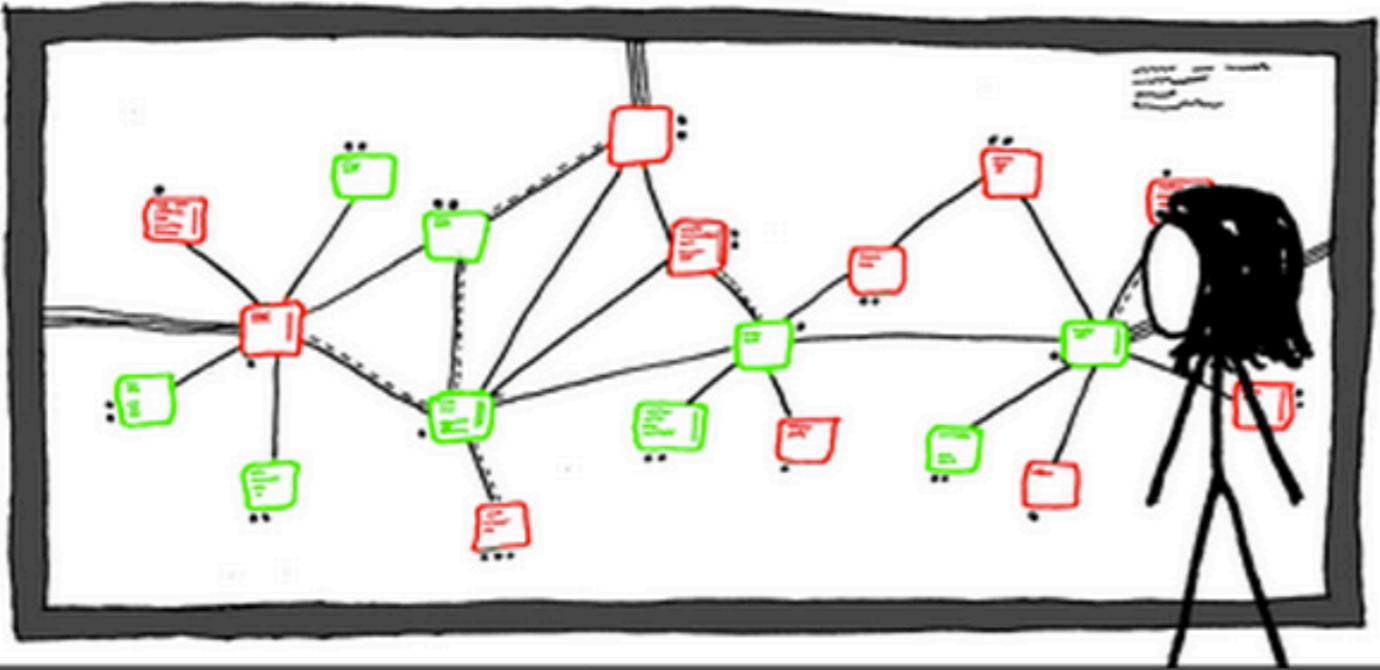
Things

Processes

Fundamentals

No-no(s)





PRETTY, ISN'T IT?

WHAT IS IT?



## net • work

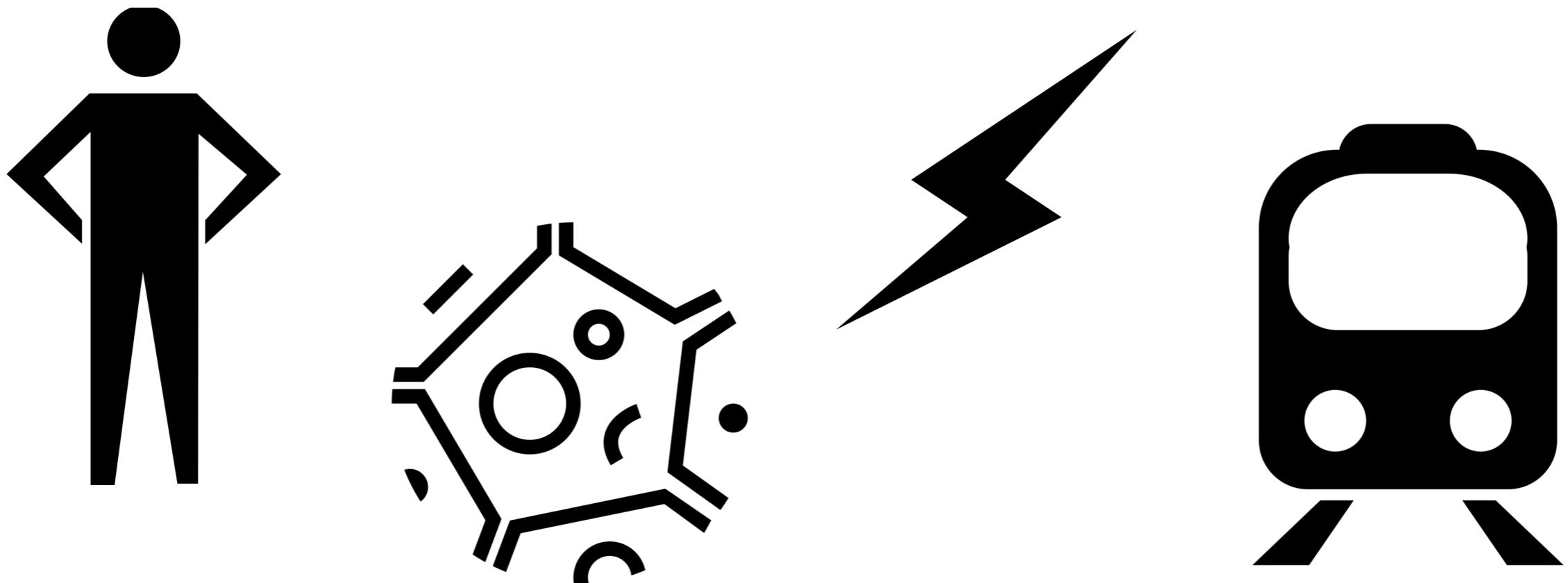
a group or system of interconnected people or things

## a • nal • y • sis

detailed examination of the elements or structure of something, typically as a basis for discussion or interpretation

# net•work a•nal•y•sis

Study of the structure of relationships between things and across things. Things include but are not limited to people, neural cells, power grids, and transportation hubs.



# questions

What does the network look like?

How connected is the network?

Which are the key entities?

Which are key subgroups?

How does network structure effect function?

# questions

- What does the network look like?
- How connected is the network?
- Which are the key entities?
- Which are key subgroups?
- How does network structure effect function?

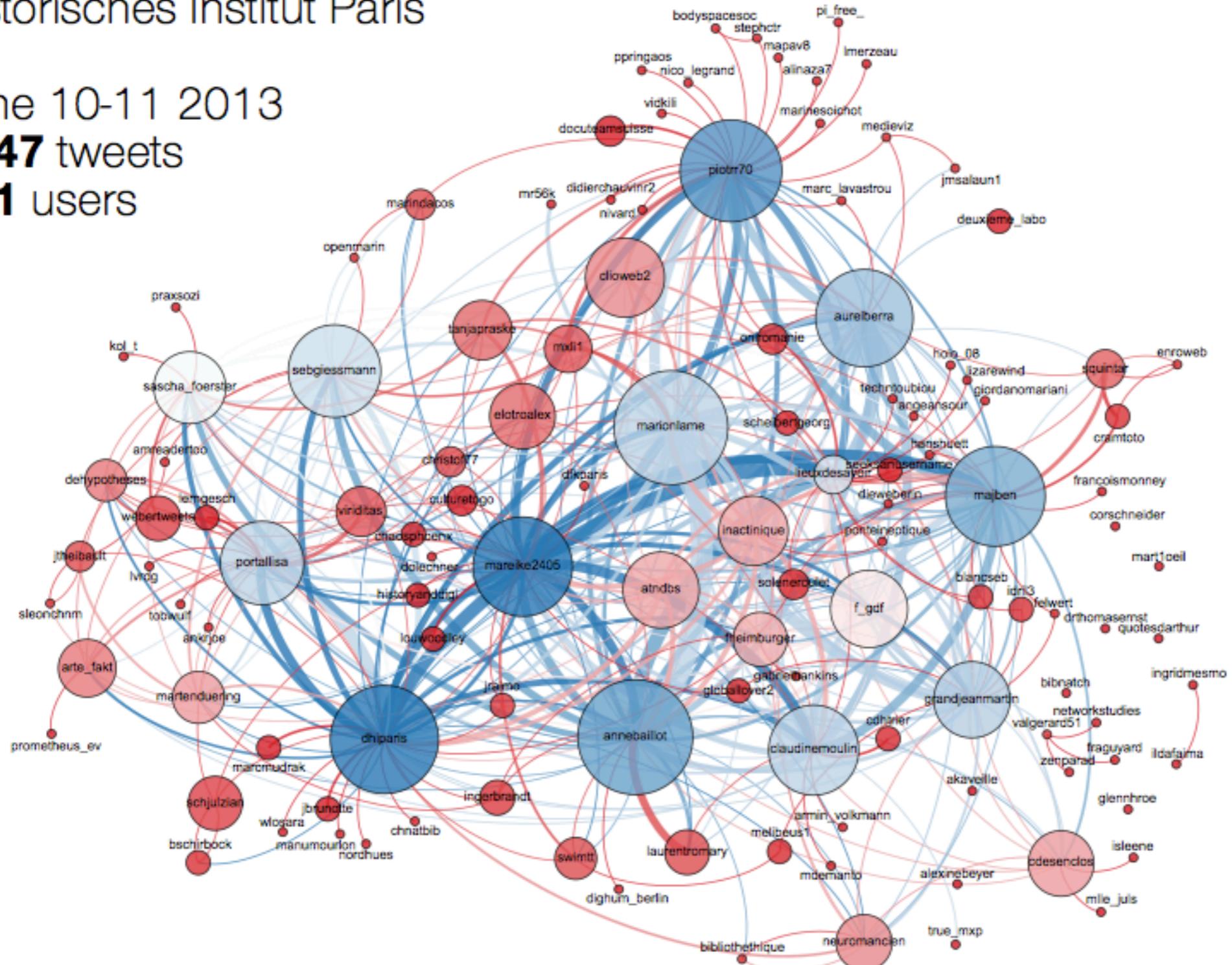
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	id_str	tweet_url	created_at	text	lang	retweet_c	screen_name	user_mentions	hashtags	query	url	in_reply_to_status_id	in_reply_to_user_id	media	lat	long			
2	44383518	https://twitter.com/dddrewd/status/44383518	2014-03-1	RT @DDDrewD:	en	3	mspalum	DD Drew	#yikes	#digitalhumanities									
3	44383461	https://twitter.com/dddrewd/status/44383461	2014-03-1	Love #digi	en	0	InfusedMedia	the_appoet	digitalhum	#digitalhum	http://www.appoet.org								
4	44383398	https://twitter.com/dddrewd/status/44383398	2014-03-1	RT @BradLis	en	1	coulterk	BradyLisaM	digitalhum	#digitalhumanities									
5	44383321	https://twitter.com/dddrewd/status/44383321	2014-03-1	RT @DDDrewD:	en	3	OkAdamO	DD Drew	#yikes	#digitalhumanities									
6	44383272	https://twitter.com/dddrewd/status/44383272	2014-03-1	WordPress	en	1	BradyLisaM		digitalhum	#digitalhumanities									
7	44383134	https://twitter.com/dddrewd/status/44383134	2014-03-1	RT @Oper	en	8	DavidMPa	OpenGLAN	digitalhum	#digitalhum	http://ow.ly/uuRmg								
8	44383098	https://twitter.com/dddrewd/status/44383098	2014-03-1	RT @Kevir	en	4	FrueheNei	KevinBGur	digitalhum	#digitalhum	http://inlustre.net/2014/03/de-commentariis-crowd-sourced-commentary-for-ancient-texts/								
9	44382725	https://twitter.com/dddrewd/status/44382725	2014-03-1	RT @DDDrewD:	en	3	fwbrantley	DD Drew	#yikes	#digitalhumanities									
10	44382710	https://twitter.com/dddrewd/status/44382710	2014-03-1	cost to ma	en	3	DDDrewDaniel		#yikes	#digitalhumanities									
11	44382695	https://twitter.com/dddrewd/status/44382695	2014-03-1	RT @Kevir	en	4	ste_tromb	KevinBGur	digitalhum	#digitalhum	http://inlustre.net/2014/03/de-commentariis-crowd-sourced-commentary-for-ancient-texts/								
12	44382000	https://twitter.com/dddrewd/status/44382000	2014-03-1	These are	en	0	karlierod	Techland	selfie	#digitalhum	http://ti.me/1e6c33h								
13	44381892	https://twitter.com/dddrewd/status/44381892	2014-03-1	RT @Kevir	en	4	electricarc	KevinBGur	digitalhum	#digitalhum	http://inlustre.net/2014/03/de-commentariis-crowd-sourced-commentary-for-ancient-texts/								
14	44381377	https://twitter.com/dddrewd/status/44381377	2014-03-1	RT @Kevir	en	4	WikiLoot	KevinBGur	digitalhum	#digitalhum	http://inlustre.net/2014/03/de-commentariis-crowd-sourced-commentary-for-ancient-texts/								
15	44379956	https://twitter.com/dddrewd/status/44379956	2014-03-1	RT @girlw	en	2	RillitoRiga	girlwithtrc	digitalhum	#digitalhum	http://www.gofundme.com/PhDigitalHumanities?pc=tw_p2								
16	44379666	https://twitter.com/dddrewd/status/44379666	2014-03-1	#StuHum	en	0	annelierug	lynnelynne	StuHum	#digitalhumanities									
17	44379612	https://twitter.com/dddrewd/status/44379612	2014-03-1	#StuHum	en	0	annelierug	lynnelynne	StuHum	#digitalhumanities									
18	44379058	https://twitter.com/dddrewd/status/44379058	2014-03-1	de comme	en	4	KevinB Gunn		digitalhum	#digitalhum	http://inlustre.net/2014/03/de-commentariis-crowd-sourced-commentary-for-ancient-texts/								
19	44378447	https://twitter.com/dddrewd/status/44378447	2014-03-1	RT @Oper	en	8	we_aRchie	OpenGLAN	digitalhum	#digitalhum	http://ow.ly/uuRmg								
20	44377926	https://twitter.com/dddrewd/status/44377926	2014-03-1	RT @DH_Cen	en	2	NicPresley	DH_Oxford	digitalhum	#digitalhum	http://dhcrowdscribe.com/crowdmap-the-crusades/								
21	44377861	https://twitter.com/dddrewd/status/44377861	2014-03-1	Am not thi	en	1	HASTACscholars		digitalhum	#digitalhumanities									
22	44377833	https://twitter.com/dddrewd/status/44377833	2014-03-1	RT @DH_Cen	en	2	ste_tromb	DH_Oxford	digitalhum	#digitalhum	http://dhcrowdscribe.com/crowdmap-the-crusades/								
23	44377815	https://twitter.com/dddrewd/status/44377815	2014-03-1	We thoug	en	2	DH_Oxford	DH_Oxford	digitalhum	#digitalhum	http://dhcrowdscribe.com/crowdmap-the-crusades/								
24	44377815	https://twitter.com/dddrewd/status/44377815	2014-03-1	We thoug	en	0	OxHumani	DH_Oxford	digitalhum	#digitalhum	http://dhcrowdscribe.com/crowdmap-the-crusades/								
25	44377798	https://twitter.com/dddrewd/status/44377798	2014-03-1	Anyone kr	en	0	_HelenaPa	kingsdh	digitalhum	#digitalhumanities									
26	44377730	https://twitter.com/dddrewd/status/44377730	2014-03-1	Hit a fundi	en	0	_HelenaPa	gofundme	DigitalHun	#digitalhum	http://www.gofundme.com/PhDigitalHumanities?pc=tw_p2_o								
27	44377723	https://twitter.com/dddrewd/status/44377723	2014-03-1	You can vo	en	0	UDMuseu	dhnnow	Editor	#digitalhum	http://ow.ly/uvljB								
28	44377723	https://twitter.com/dddrewd/status/44377723	2014-03-1	You can vo	en	0	Tracy_Jent	dhnnow	Editor	#digitalhum	http://ow.ly/uvlbR								
29	44377519	https://twitter.com/dddrewd/status/44377519	2014-03-1	RT @girlw	en	2	CSHoggard	girlwithtrc	digitalhum	#digitalhum	http://www.gofundme.com/PhDigitalHumanities?pc=tw_p2								
30	44377467	https://twitter.com/dddrewd/status/44377467	2014-03-1	My talents	en	2	girlwithtrc	_HelenaPa	digitalhum	#digitalhum	http://www.gofundme.com/PhDigitalHumanities?pc=tw_p2								
31	44377145	https://twitter.com/dddrewd/status/44377145	2014-03-1	RT @dhno	en	2	andrea_vil	dhnnow	digitalhum	#digitalhumanities									
32	44377034	https://twitter.com/dddrewd/status/44377034	2014-03-1	RT @dhno	en	2	Asteria_sr	dhnnow	digitalhum	#digitalhumanities									
33	44376967	https://twitter.com/dddrewd/status/44376967	2014-03-1	RT @kram	en	3	DavidUnde	kramermj	twittersto	#digitalhum	http://hnn.us/article/154929								
34	44376890	https://twitter.com/dddrewd/status/44376890	2014-03-1	RT @kram	en	3	lljohnston	kramermj	twittersto	#digitalhum	http://hnn.us/article/154929								
35	44376422	https://twitter.com/dddrewd/status/44376422	2014-03-1	@IreRubin	en	2	dhnnow	IreRubino	digitalhum	#digitalhumanities	IreRubino	4.44E+17							
36	44376216	https://twitter.com/dddrewd/status/44376216	2014-03-1	RT @kram	en	3	MakDigHi	kramermj	twittersto	#digitalhum	http://hnn.us/article/154929								
37	44376188	https://twitter.com/dddrewd/status/44376188	2014-03-1	@wthom	en	3	kramermj	wthomas	twittersto	#digitalhum	http://hnn.us/article/154929								

# Digital Humanities at Deutsches Historisches Institut Paris

June 10-11 2013

**1147** tweets

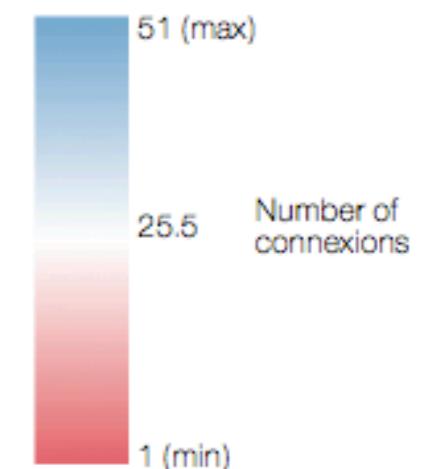
**121** users



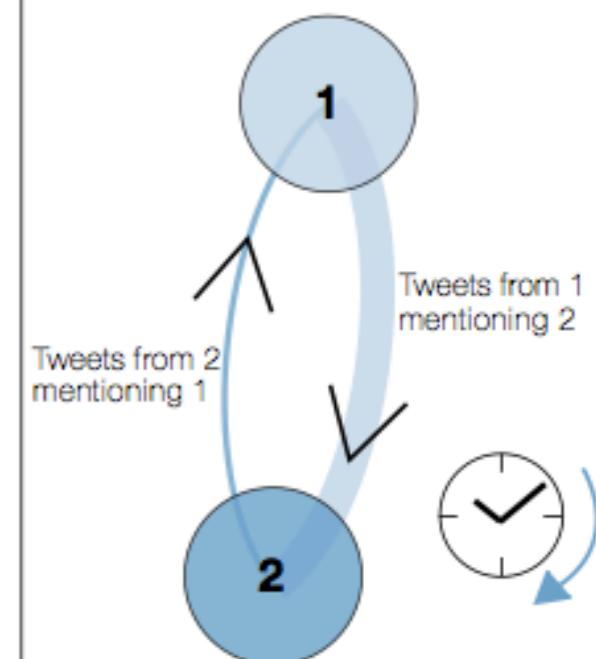
**NUMBER OF TWEETS WRITTEN**



**DEGREE CENTRALITY**



**EDGES**



# questions

- What does the network look like?
- How connected is the network?
- Which are the key entities?
- Which are key subgroups?
- How does network structure effect function?

A Support Network has a major group centered on a hub account, in this case DellCares , which connects to many spokes who each get a reply from the hub but do not connect to each other.

Isolates - users who have not been engaged with the support account (or any other account) even though they used the keyword phrases. .

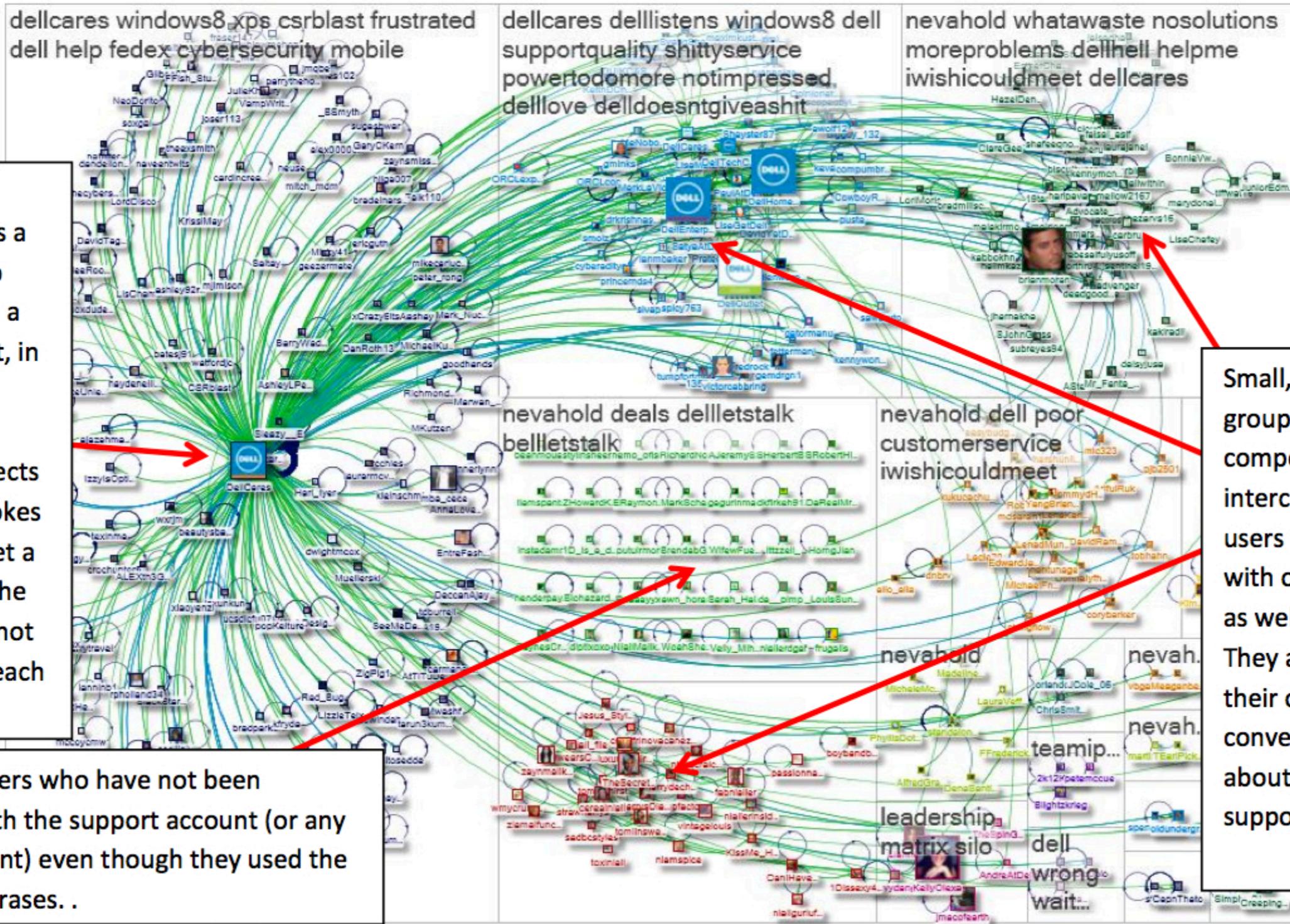


Figure 44: Network graph of 388 Twitter users whose tweets contained dellistens OR dellcares collected between February 12-19, 2013. There is a green line or “edge” for each follows relationship. There is a blue edge for each “replies-to” and “mentions” relationship in a tweet. There is a self-loop edge for each tweet that is not a “replies-to” or “mentions.”

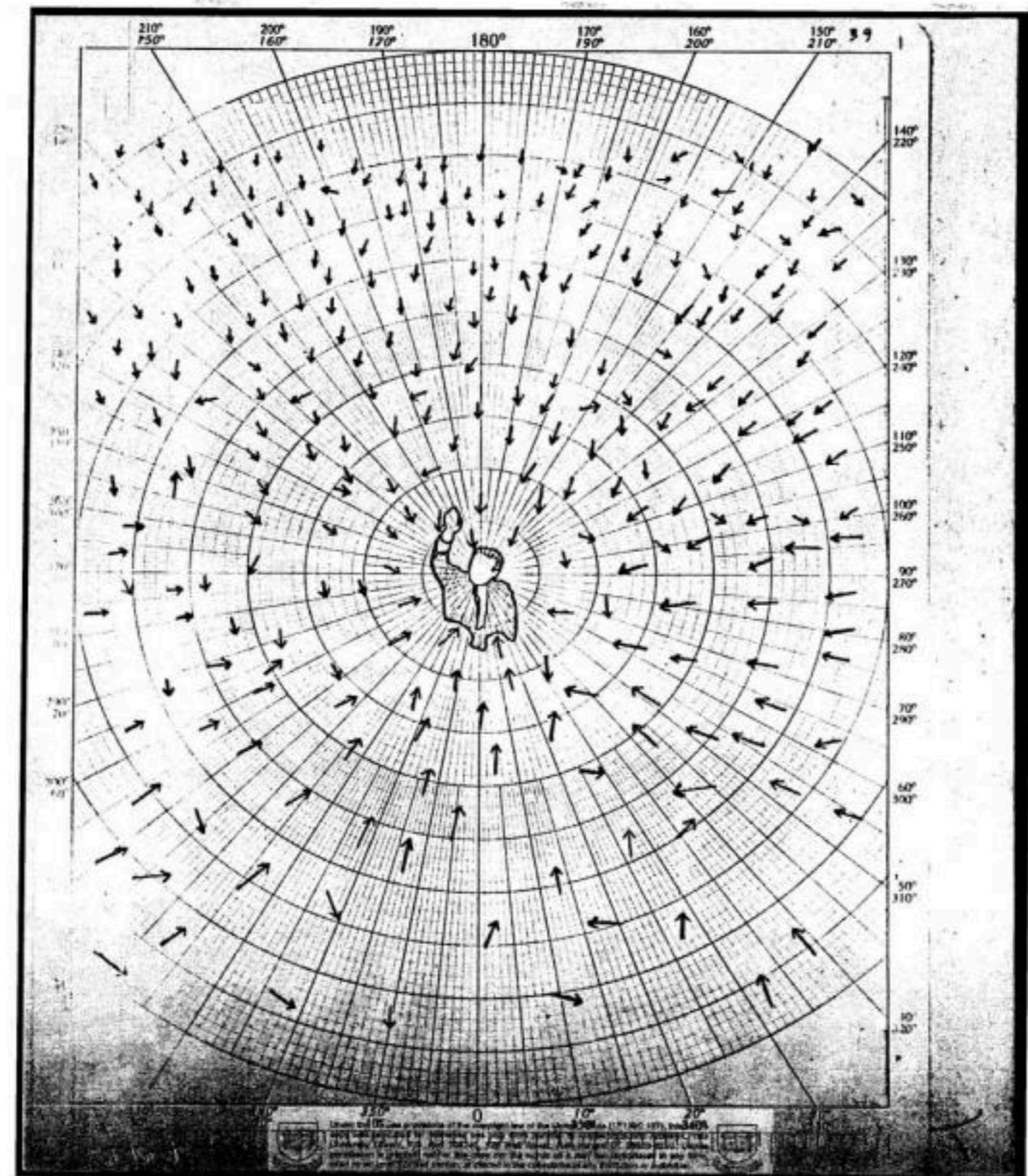
# thought experiment

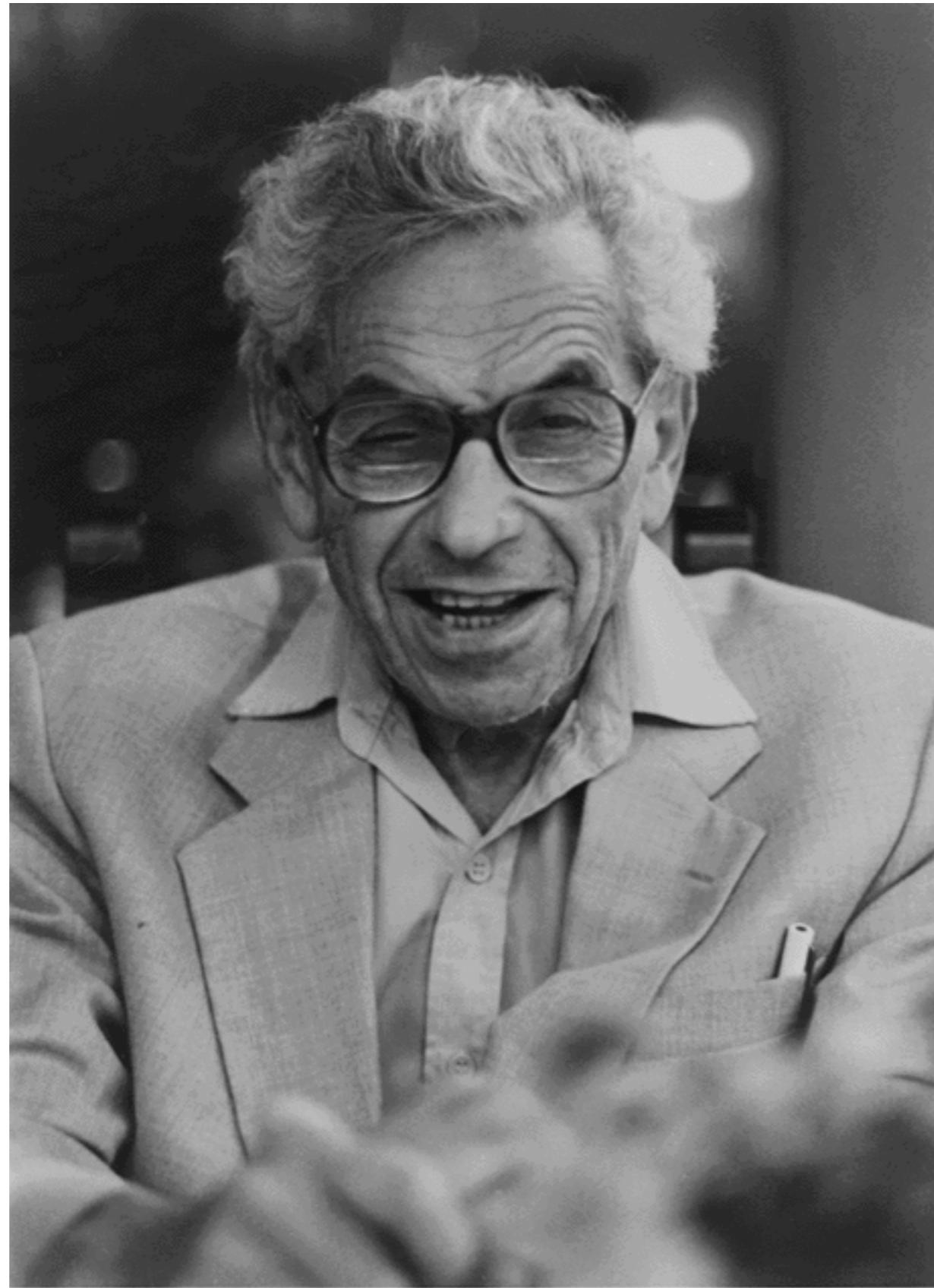
the enemy of my enemy is my friend  
the friend of my friend is my friend  
the friend of my enemy is my enemy  
the enemy of my friend is my enemy

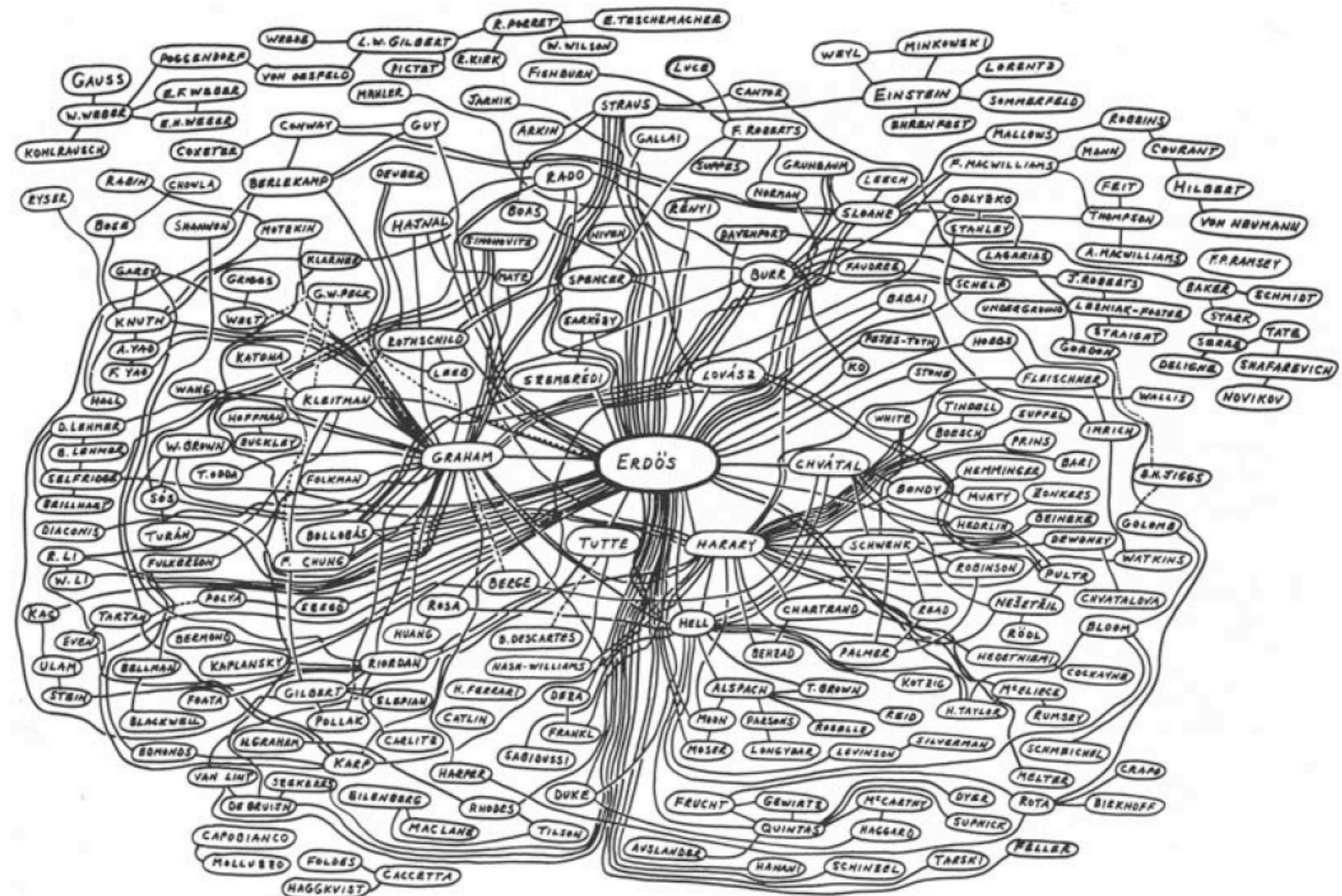
*I read somewhere that  
everybody on this planet is  
separated by only six other  
people. Six degrees of  
separation. Between us and  
everyone else on the planet.*

John Guare, *Six Degrees of Separation*











**KEVIN  
BACON**



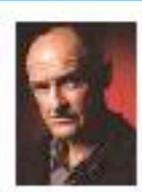
*Yunjin Kim*  
*Iron Palm(2002)*



*Rodrigo Santoro*  
*Love Actually(2003)*



*Naveen Andrews*  
*The English Patient(1996)*



*Terry O'Quinn*  
*Mrs. Soffel(1984)*



*Maggie Grace*  
*The Fog(2005)*



*Denzel Washington*  
*L. Scott Caldwell*  
*Devil in a Blue Dress(1995)*



*Malcolm David Kelley*  
*Antwon Fischer(2002)*



*Daniel Dae Kim*  
*The Jackal(1997)*



*Dominic Monaghan*  
*LOTR: ROTK (2003)*



*Harold Perrineau*  
*Macbeth in Manhattan(1999)*



*Kim Delgado*  
*Patriot Games(1992)*



*Colin Firth* Maury Chakin Sonja Bennet Beau Starr David Hayman  
**Where the Truth Lies(2005)**



**Kevin Bacon**



Michael Emerson  
*The Imposters(1998)*



Oliver Platt John Elson John Lafayette  
**Loverboy(2005)**



Demi Moore  
**A Few Good Men(1992)**



No Image Jon Smith Ed O'Keefe Adam Nelson  
**Mystic River(2003)**



*Adewale Akinnuoye-Agbaje*  
*Legionnaire(1998)*



*Evangeline Lily*  
*The Long Weekend(2005)*



*Kiele Sanchez*  
*Stuck on You(2003)*



*Josh Holloway*  
*Dr. Benny(2003)*



*Michelle Rodriguez*  
*S.W.A.T.(2003)*



*Kathleen Quinlan*  
*Apollo 13(1995)*



*Paul Dooly*  
*Telling Lies in America(1997)*



*Kelly Bishop*  
*Queens Logic(1991)*



*Steve Martin*  
*Novocaine(2001)*



*Sherri Shepard* Octavia Spencer  
**Beauty Shop(2005)**



*Matthew Fox*  
*My Boyfriend's Back(1993)*



*Emilie de Ravin*  
*The Hills Have Eyes(2006)*



*Cynthia Watros*  
*Cafe Society(1995)*



*Sam Anderson*  
*Movers and Shakers(1985)*



*Jorge Garcia*  
*King of the Open Mic's*



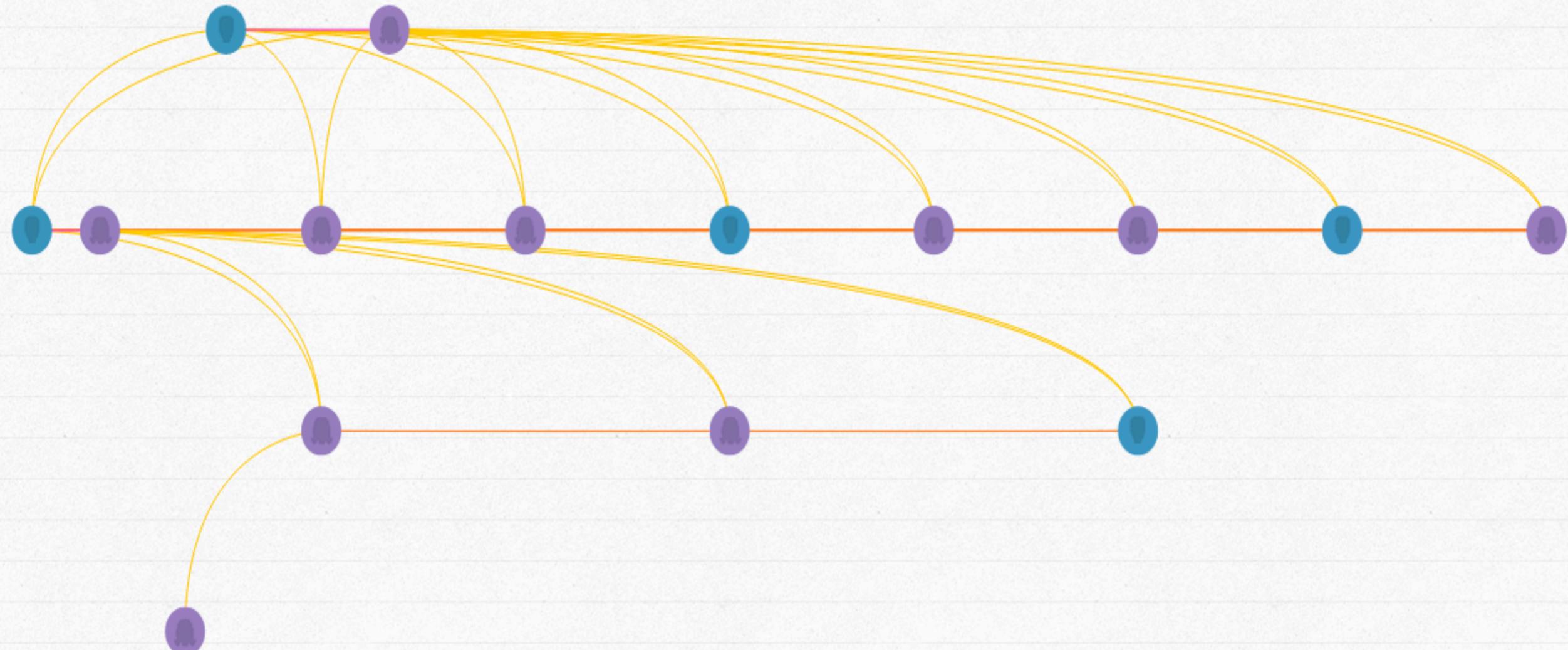
*Ian Somerhalder*  
*Pulse(2006)*

William Shakespeare  
1564 - 1616



*The family of William Shakespeare*

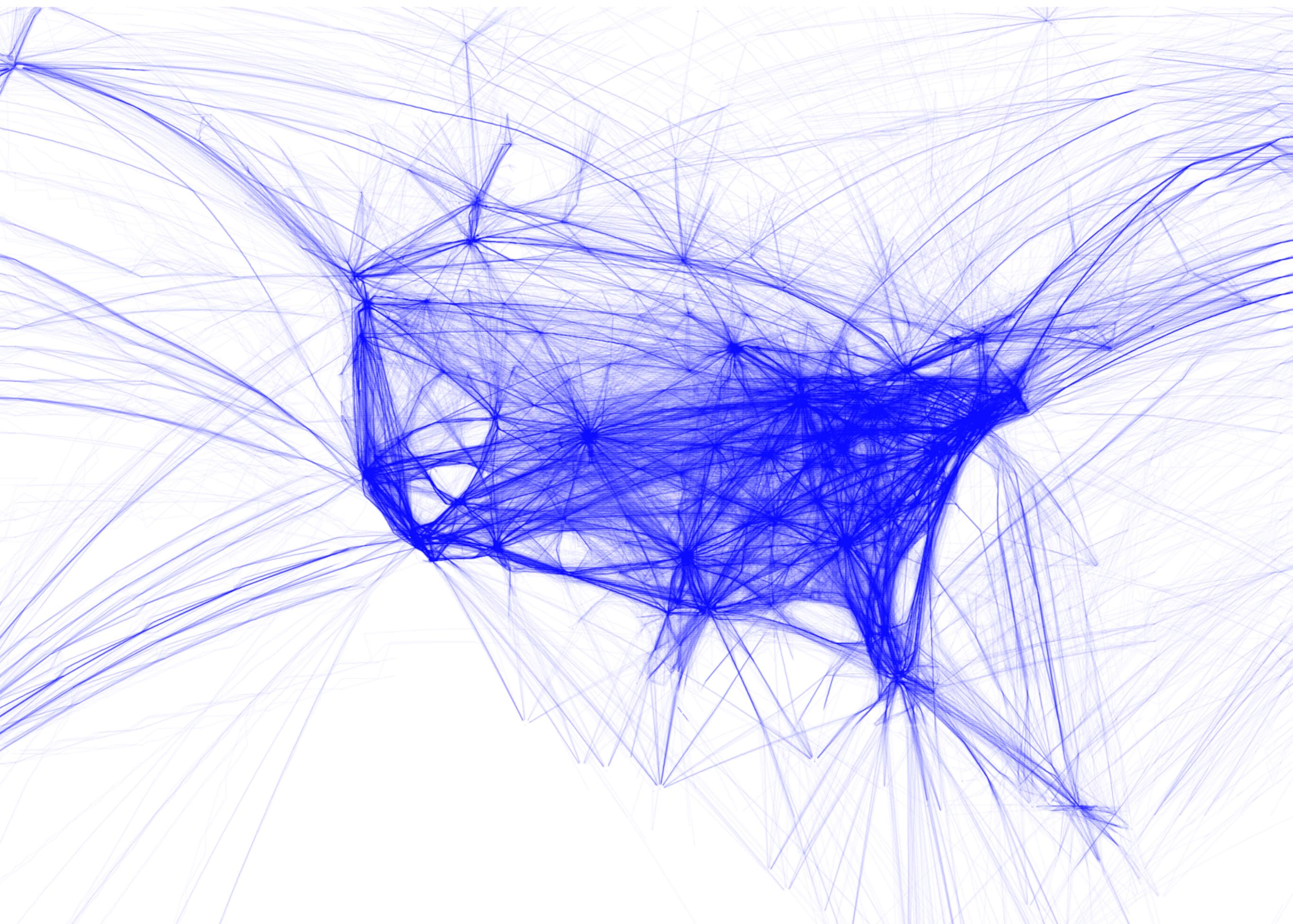
NETWORK





*A relational view of reality, not a separate part of reality called networks*

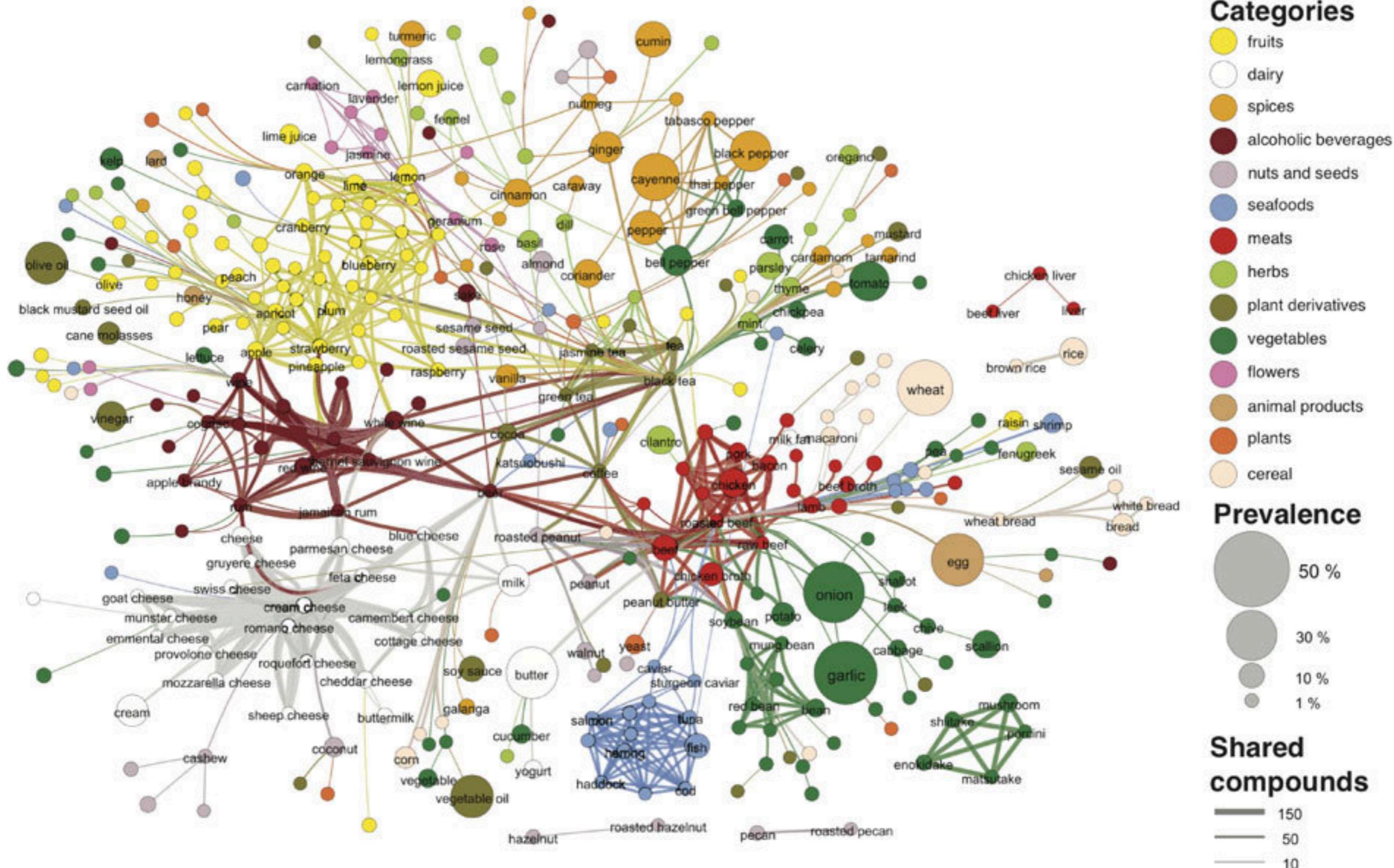
Claire LeMercier, “Formal Network Methods in History: Why and How?”



We introduce a flavor network that captures the flavor compounds shared by culinary ingredients.

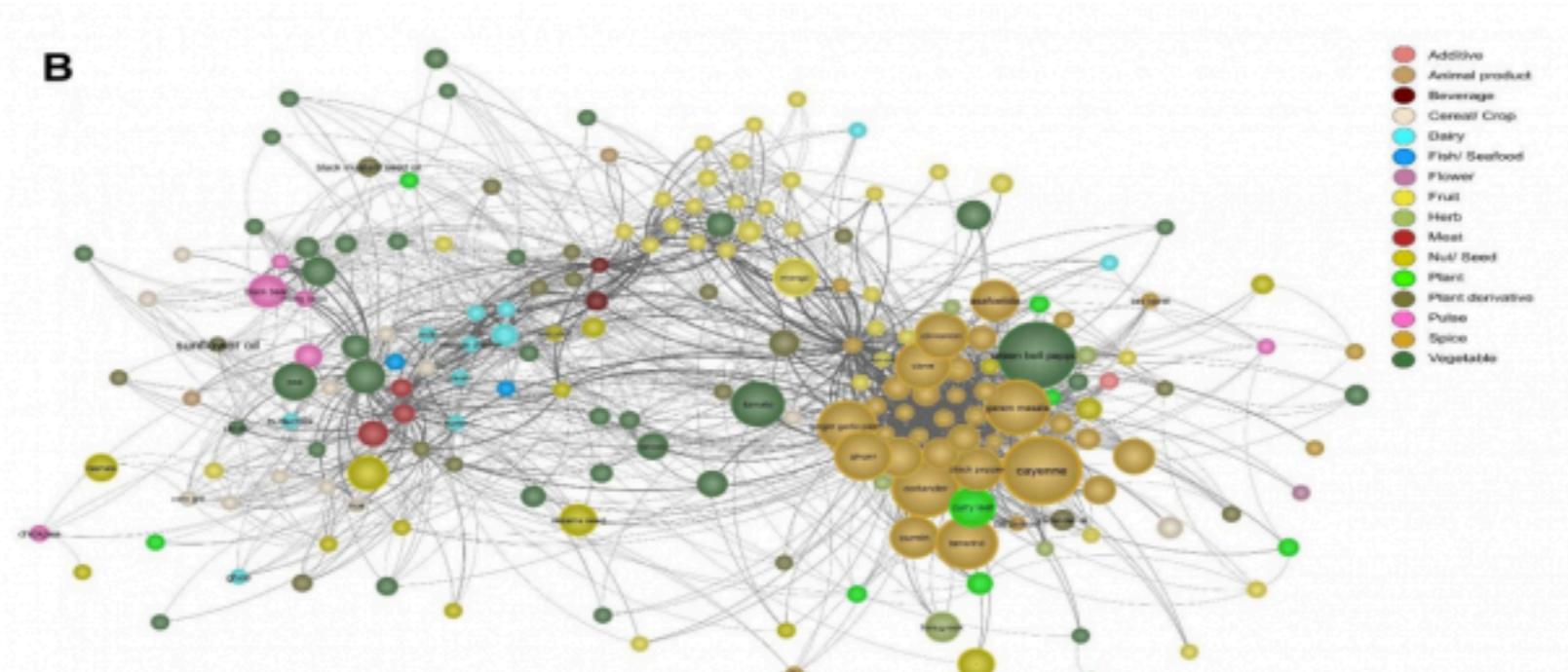
Western cuisines show a tendency to use ingredient pairs that share many flavor compounds, supporting the so-called food pairing hypothesis. By contrast, East Asian cuisines tend to avoid compound sharing ingredients. Given the increasing availability of information on food preparation, our data-driven investigation opens new avenues towards a systematic understanding of culinary practice.

Yong-Yeol Ahn, Sebastian E. Ahnert, James P. Bagrow & Albert-László Barabási, “Flavor network and the principles of food pairing”, Nature, 2011

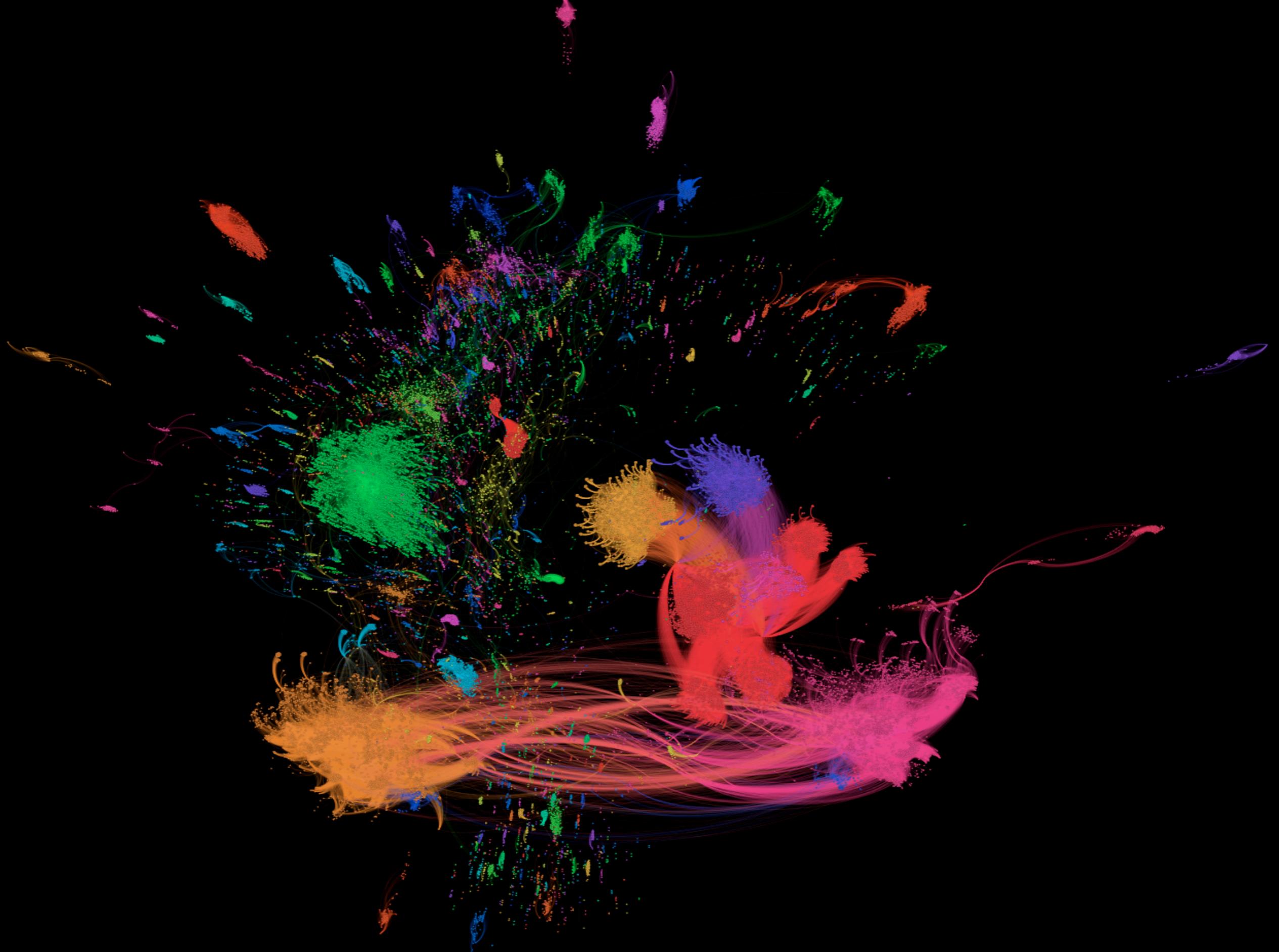


# Data Mining Indian Recipes Reveals New Food Pairing Phenomenon

By studying the network of links between Indian recipes, computer scientists have discovered that the presence of certain spices makes a meal much less likely to contain ingredients with flavors in common.







Ian Milligan, Using WarcBase to Generate a Linkgraph of the Wide Web Scrape  
<http://ianmilligan.ca/2015/03/03/using-warcbase-to-generate-a-link-graph-of-the-wide-web-scrape/>

# basic process

Specify question

Specify nodes, edges, network boundaries

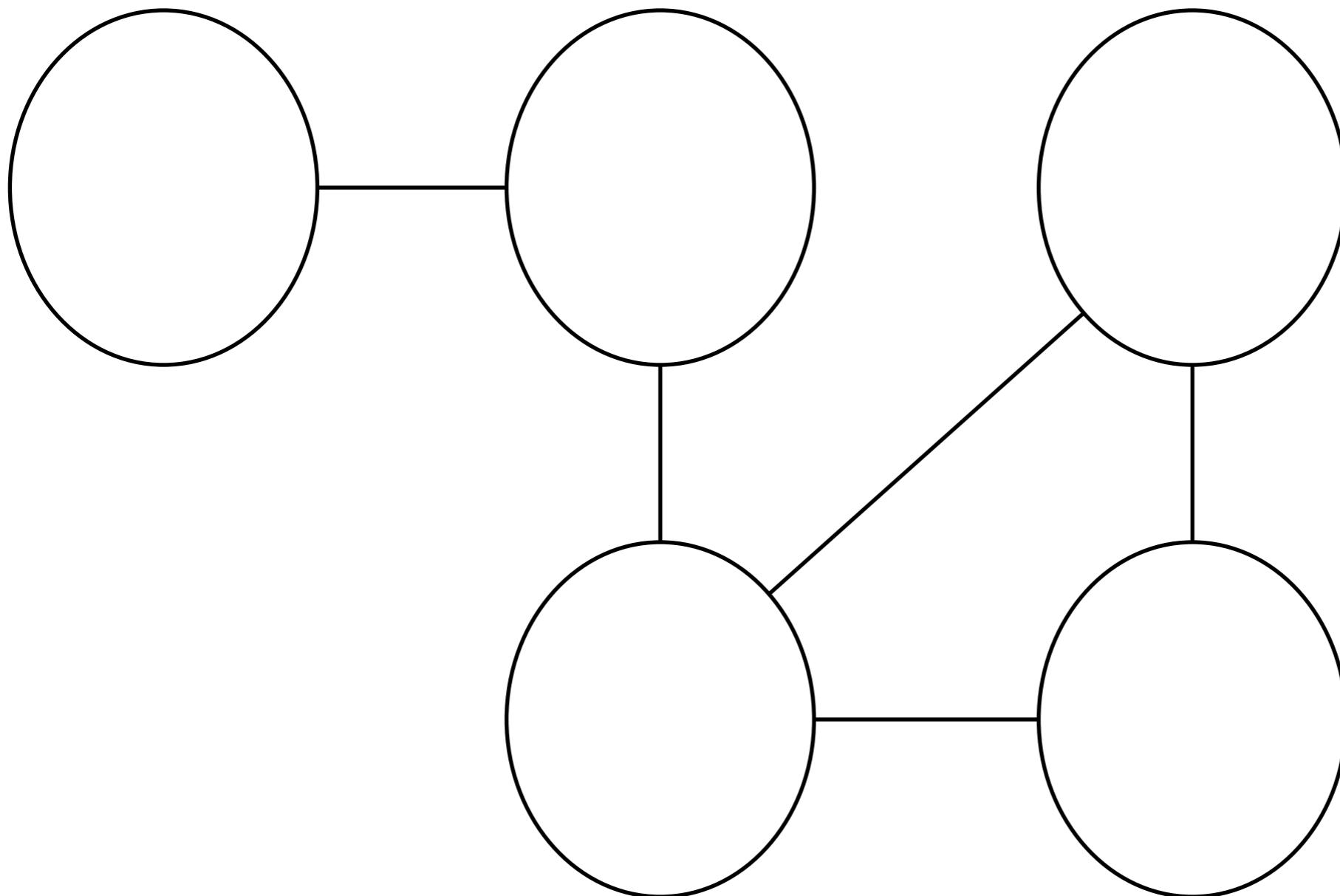
Find or create relational data

Analyze relational data

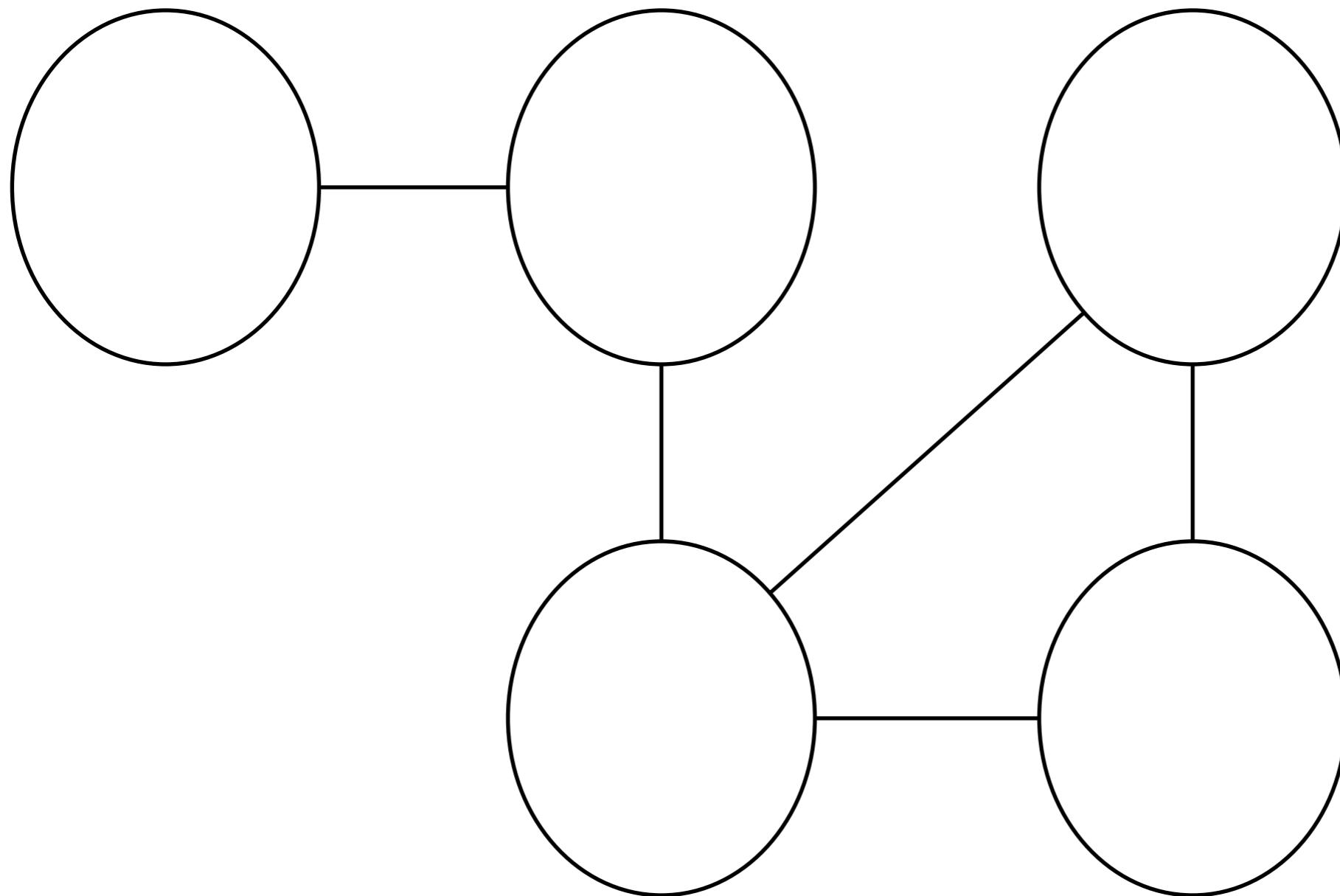
Interpret results

Repeat

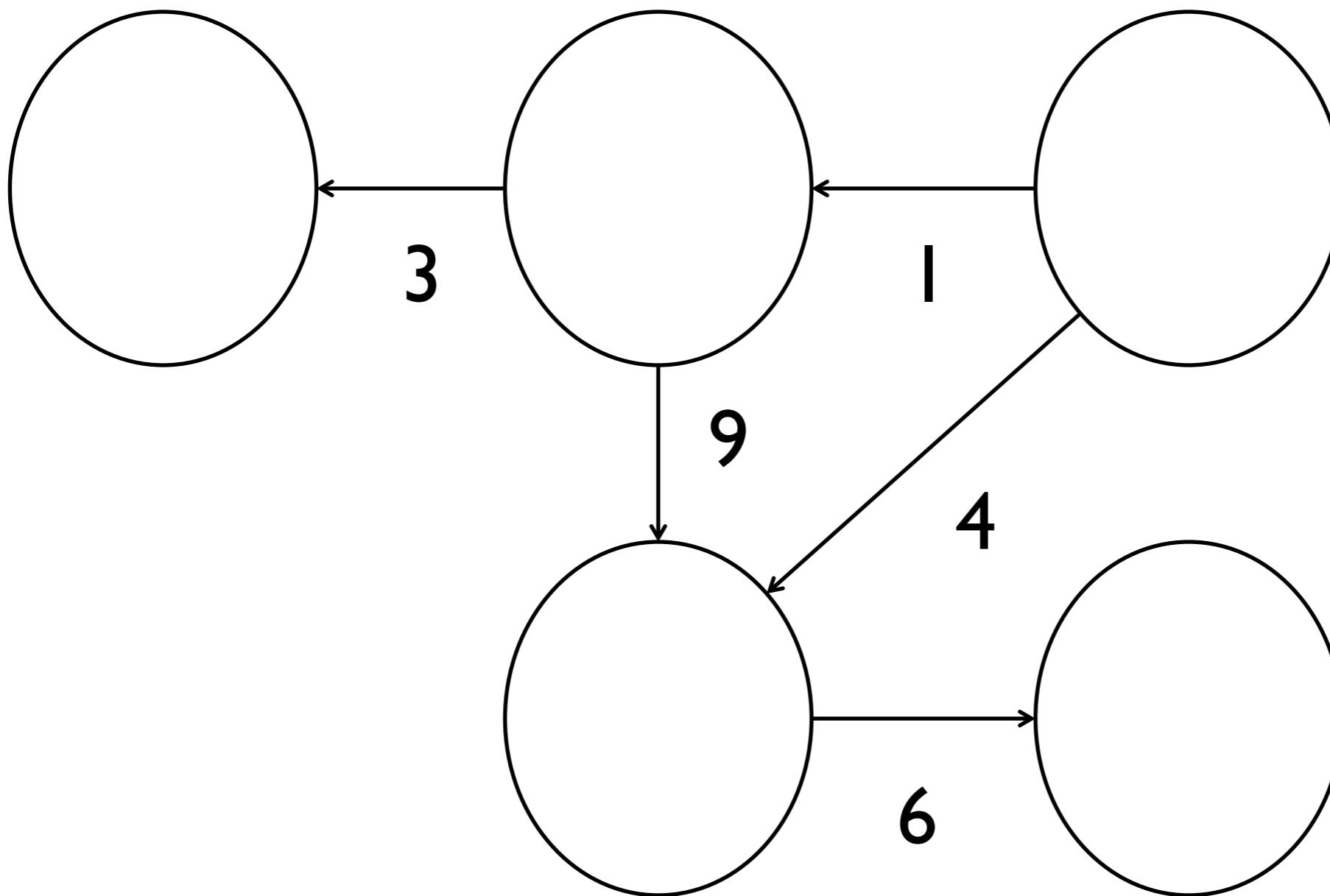
# Graphs



nodes & edges

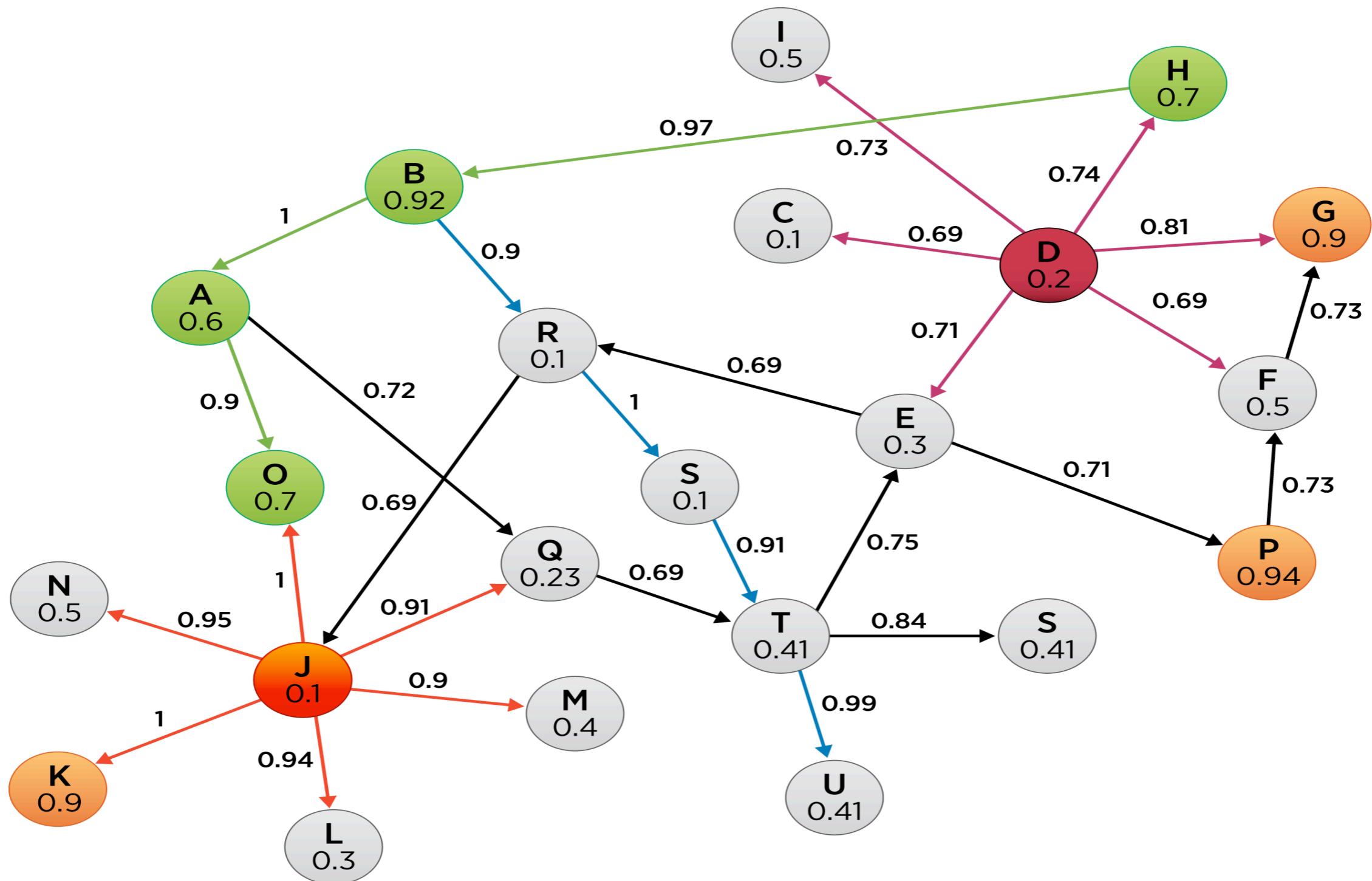


directed



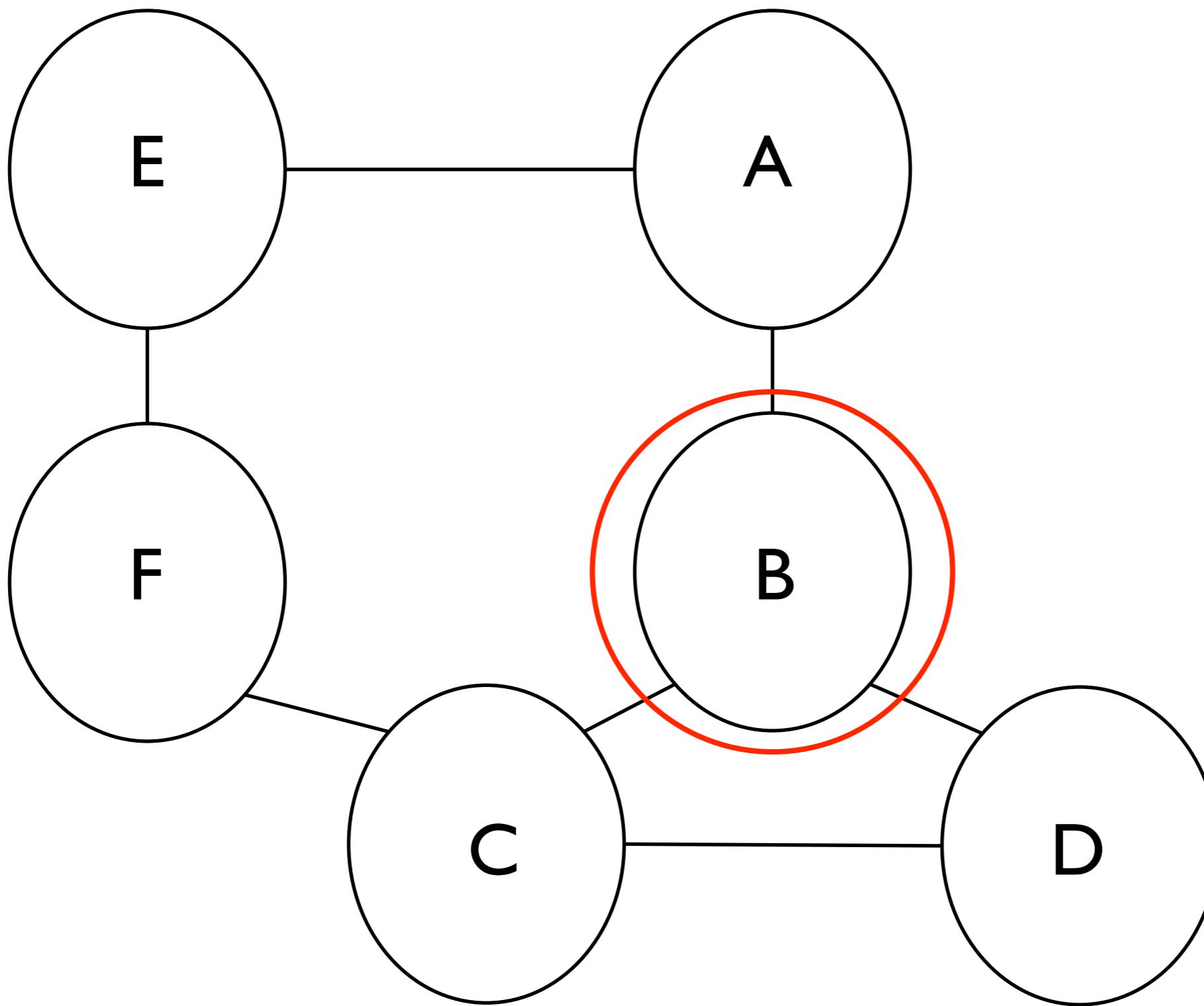
# path

a sequence of nodes, where each node in the sequence is connected by an edge

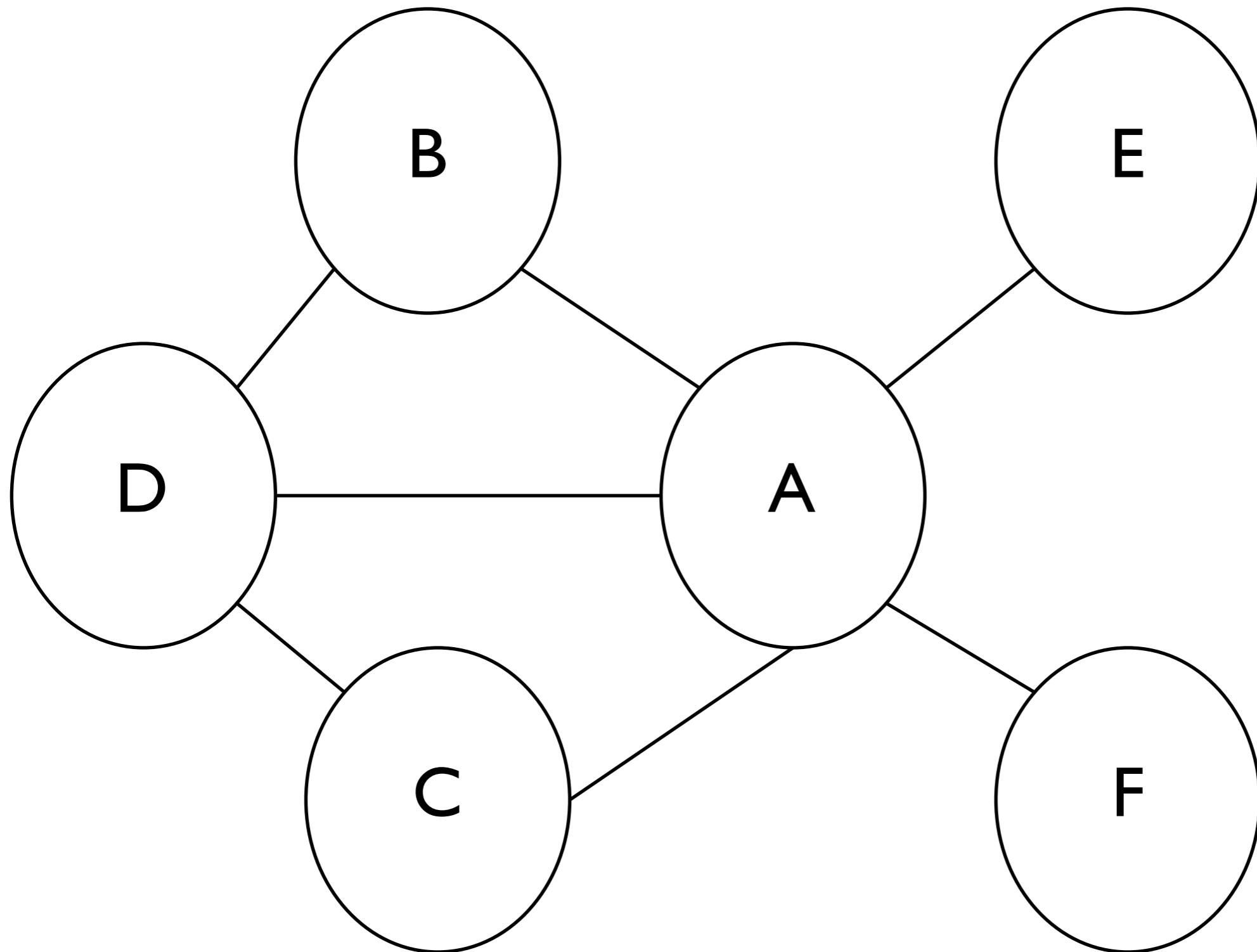


# pivotal node

node is pivotal if it lies on shortest path between two pairs of nodes

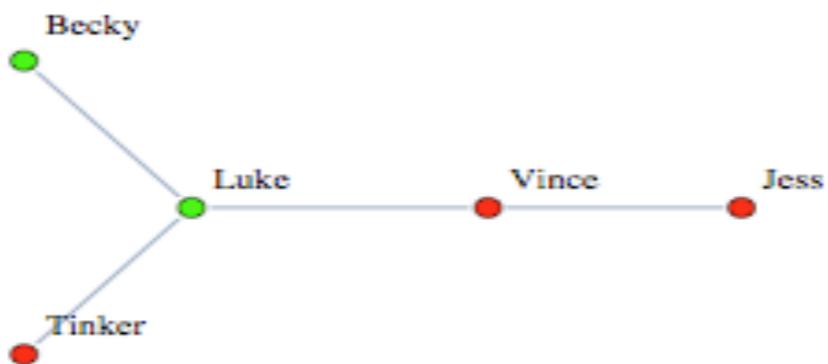
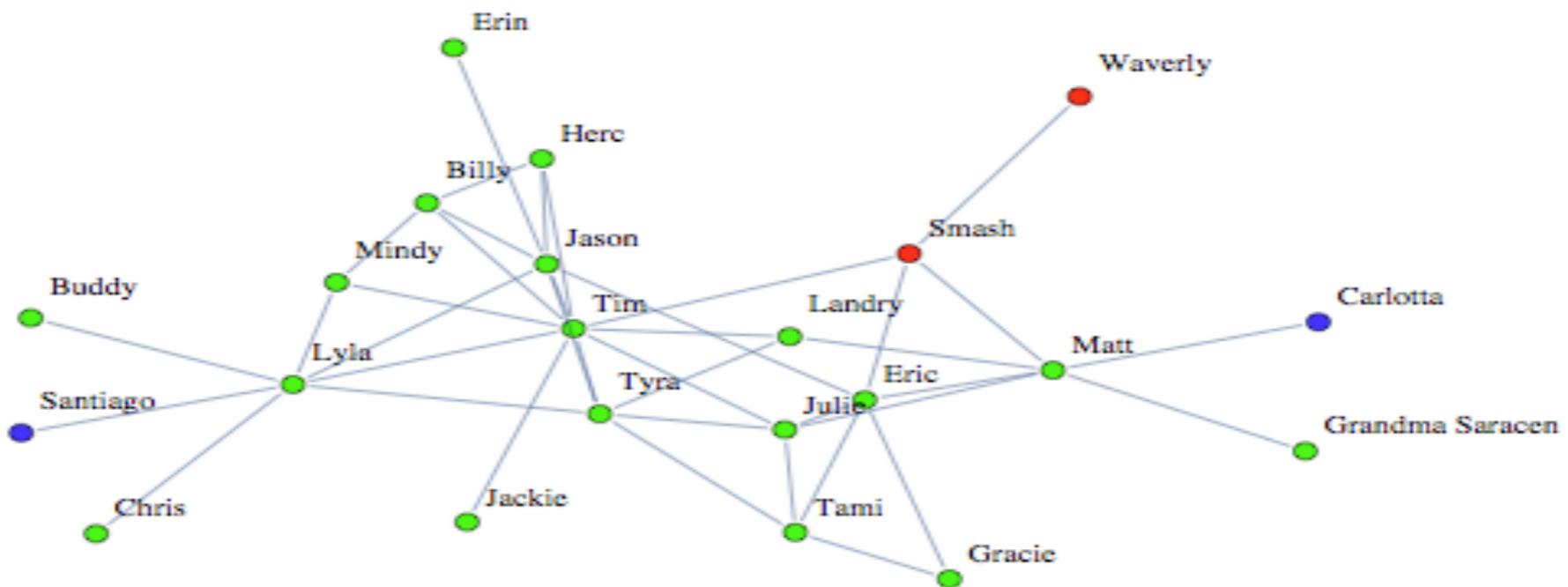


# gatekeeping



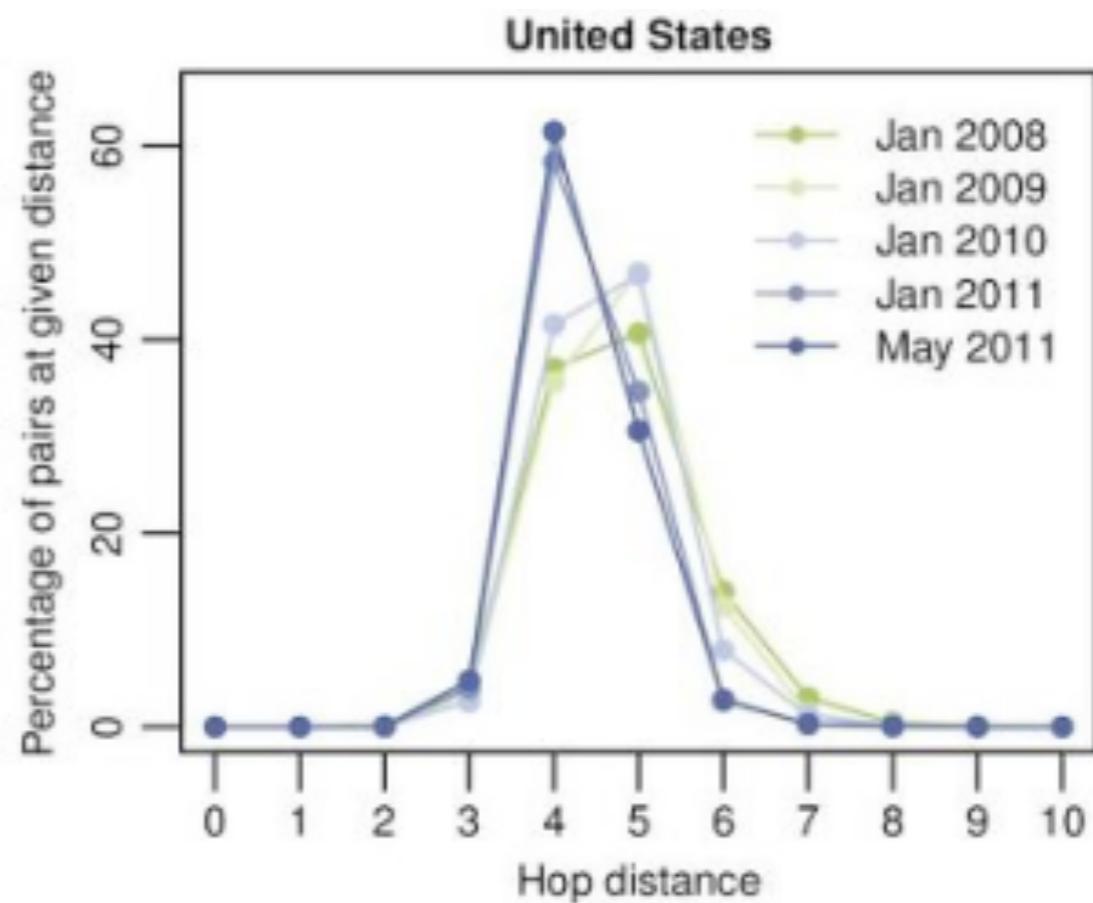
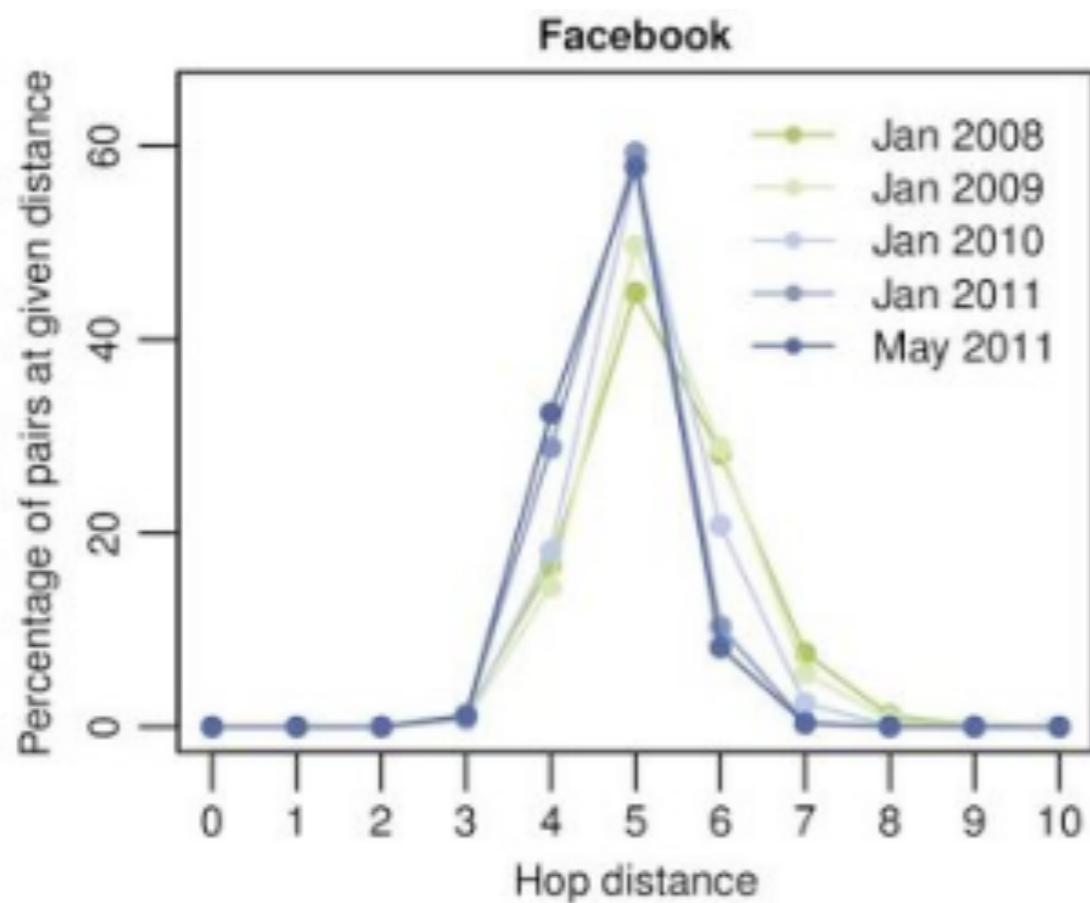
# component

a subset of connected nodes



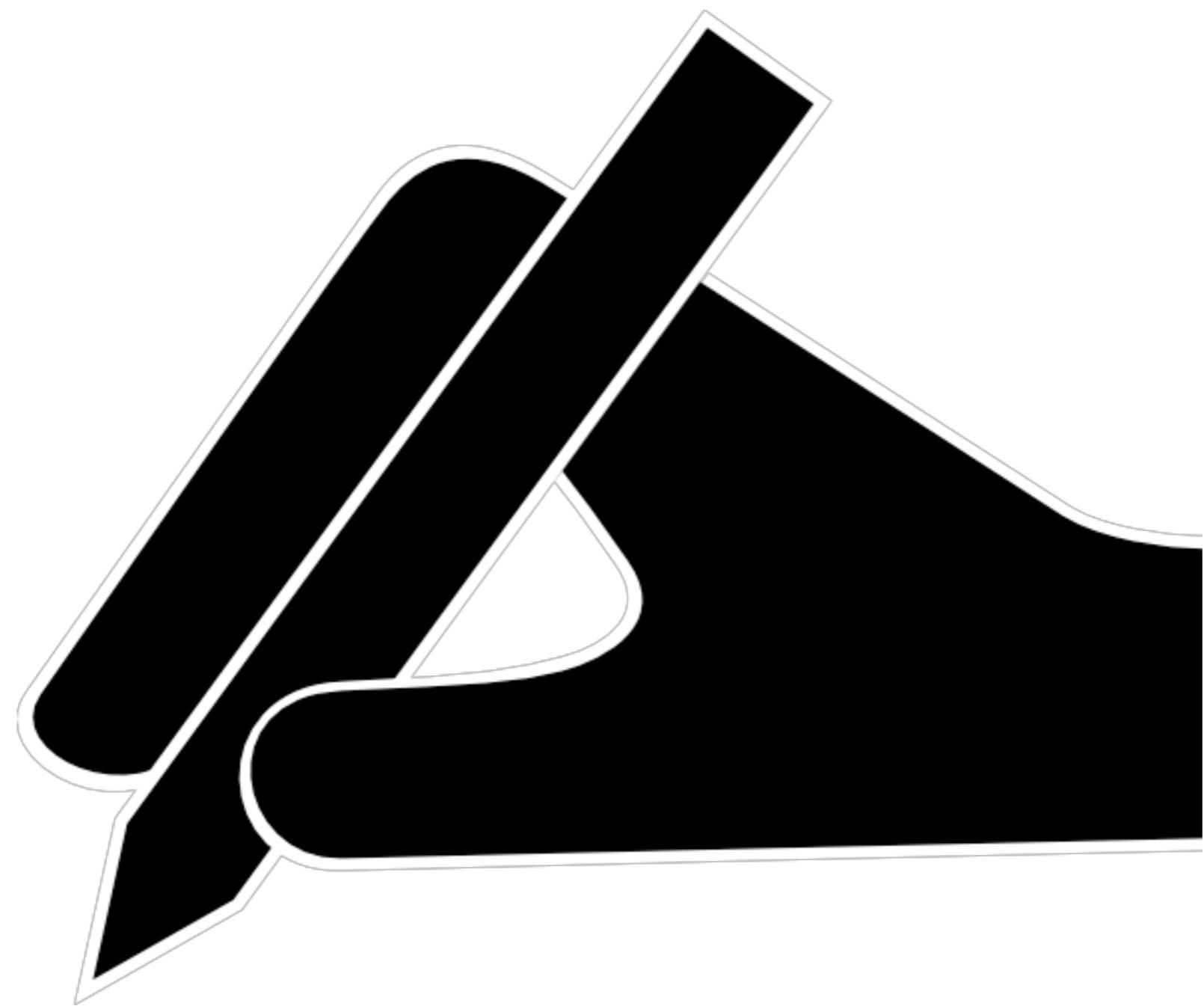
# giant component

a component that contains the majority of available nodes



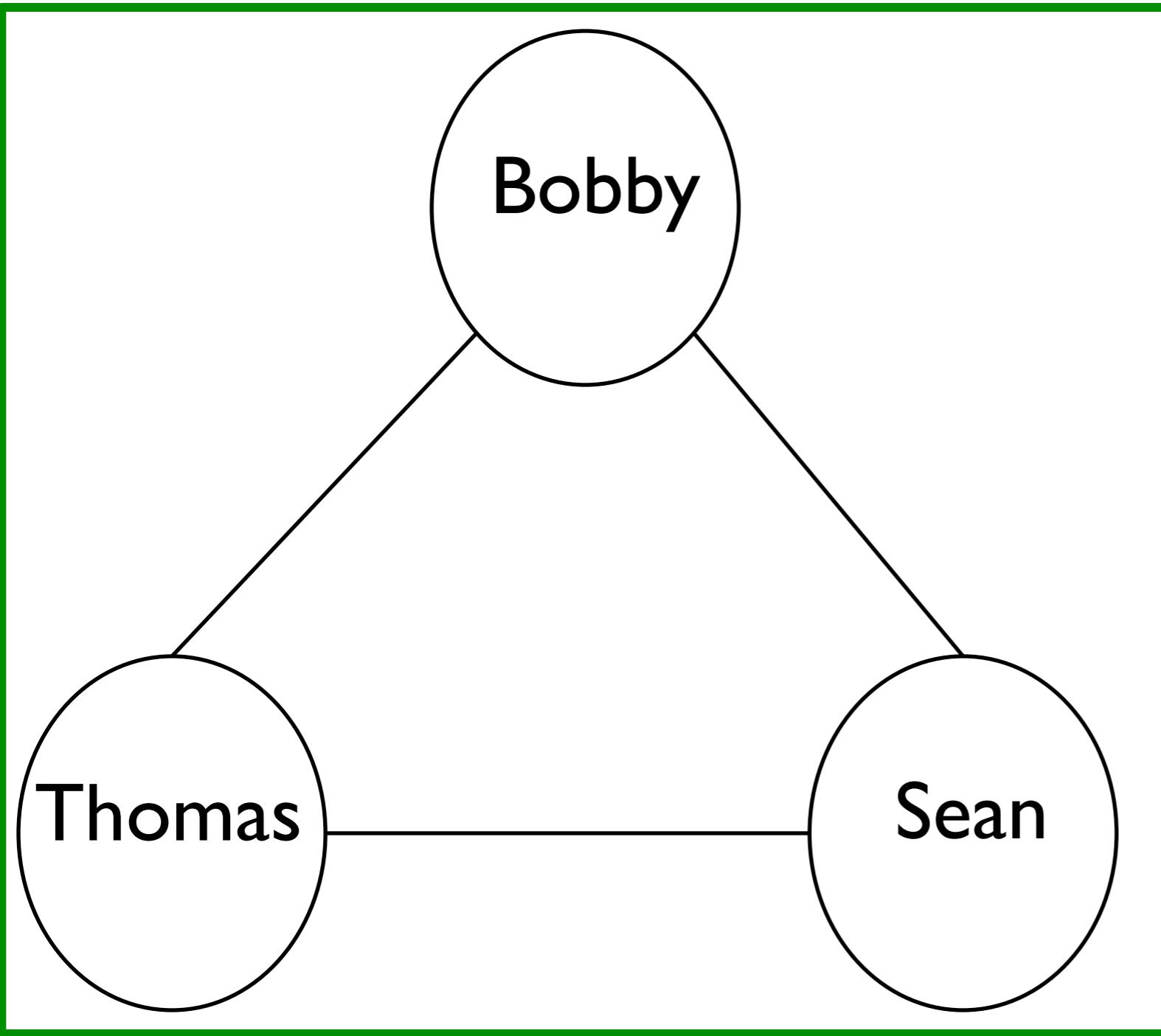
# thought experiment

the enemy of my enemy is my friend  
the friend of my friend is my friend  
the friend of my enemy is my enemy  
the enemy of my friend is my enemy



The enemy of my enemy is my friend  
The friend of my friend is my friend  
The friend of my enemy is my enemy  
The enemy of my friend is my enemy

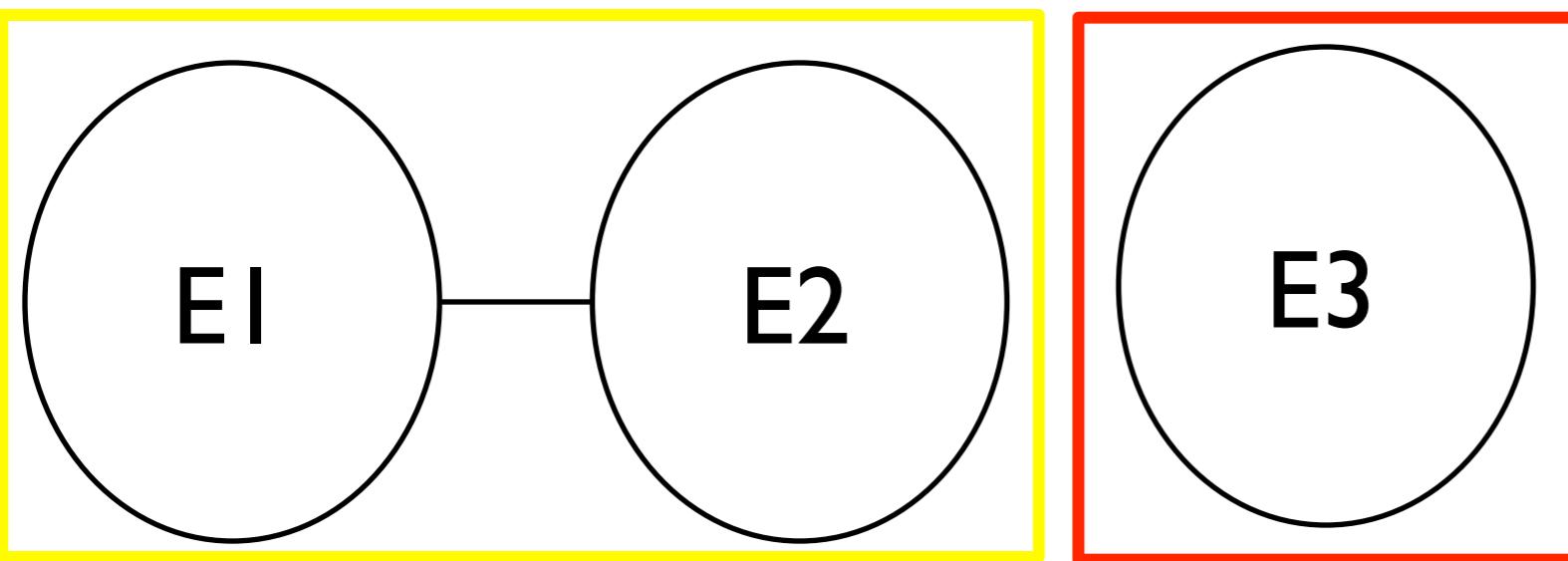
	T	B	S	E1	E2	E3
T	0	I	I	0	0	0
B	I	0	I	0	0	0
S	I	I	0	0	0	0
E1	0	0	0	0	I	0
E2	0	0	0	I	0	0
E3	0	0	0	0	0	0



**triad** – group of three dyads

**dyad** – pair of nodes

**isolate** – unconnected node

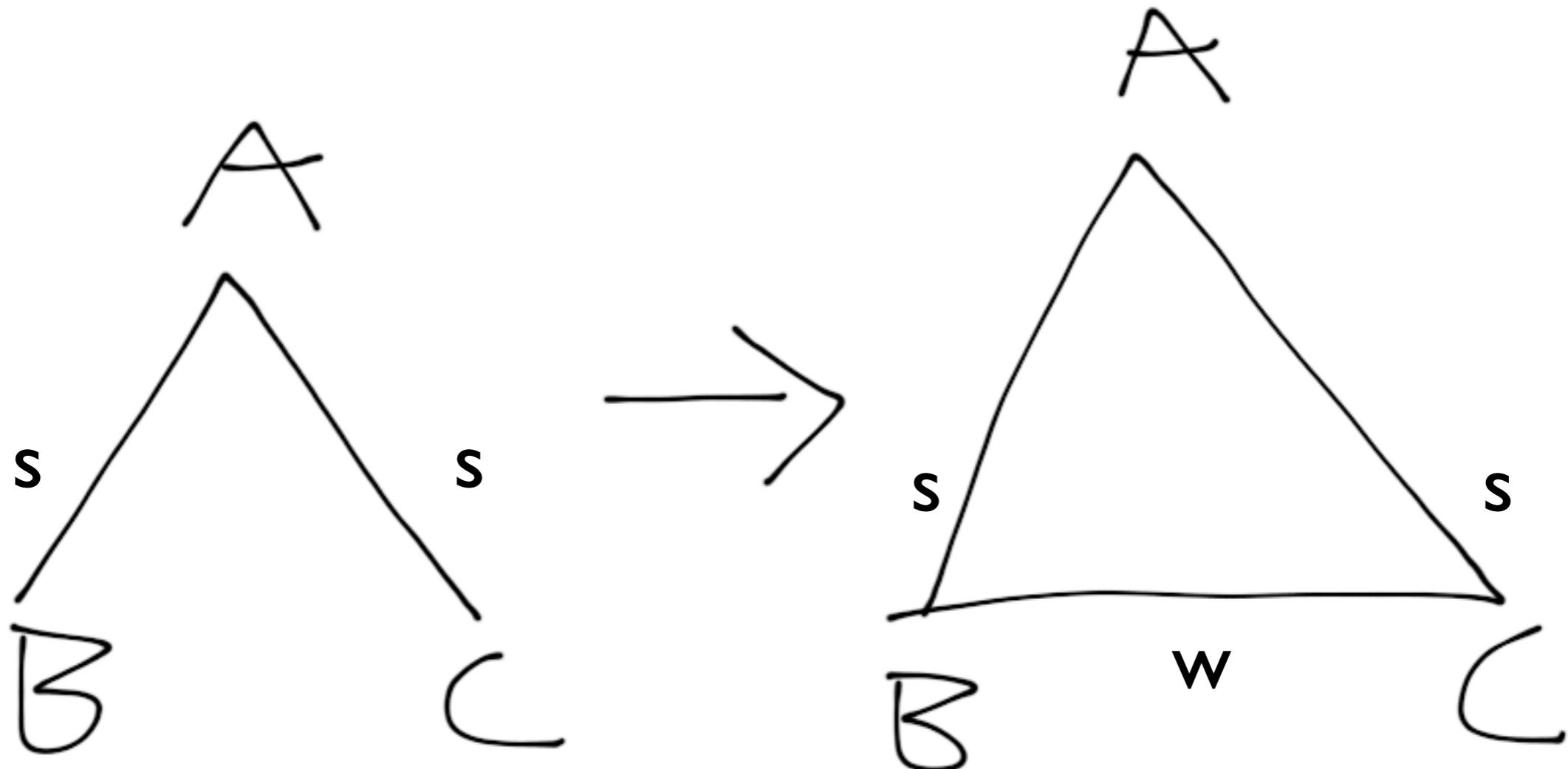


*One of the powerful roles  
that networks play is to  
bridge the local and global  
- to offer explanations for  
how simple processes at the  
level of individual nodes  
can have complex effects  
that ripple through the  
population as a whole...*

David Easley & John Kleinberg

# triadic closure

if two nodes share relation, likelihood is increased that all nodes will be connected

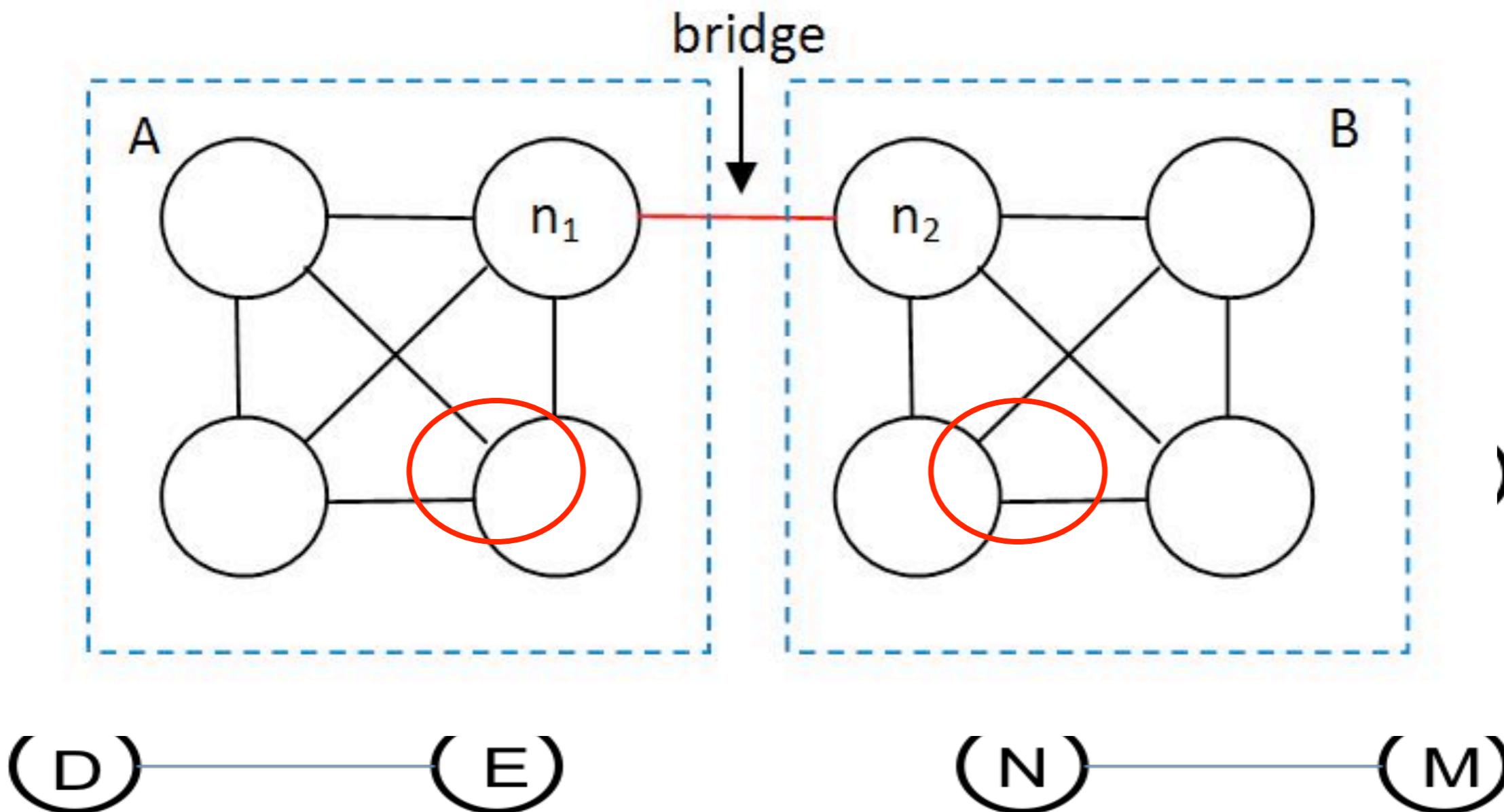


# the strength of weak ties



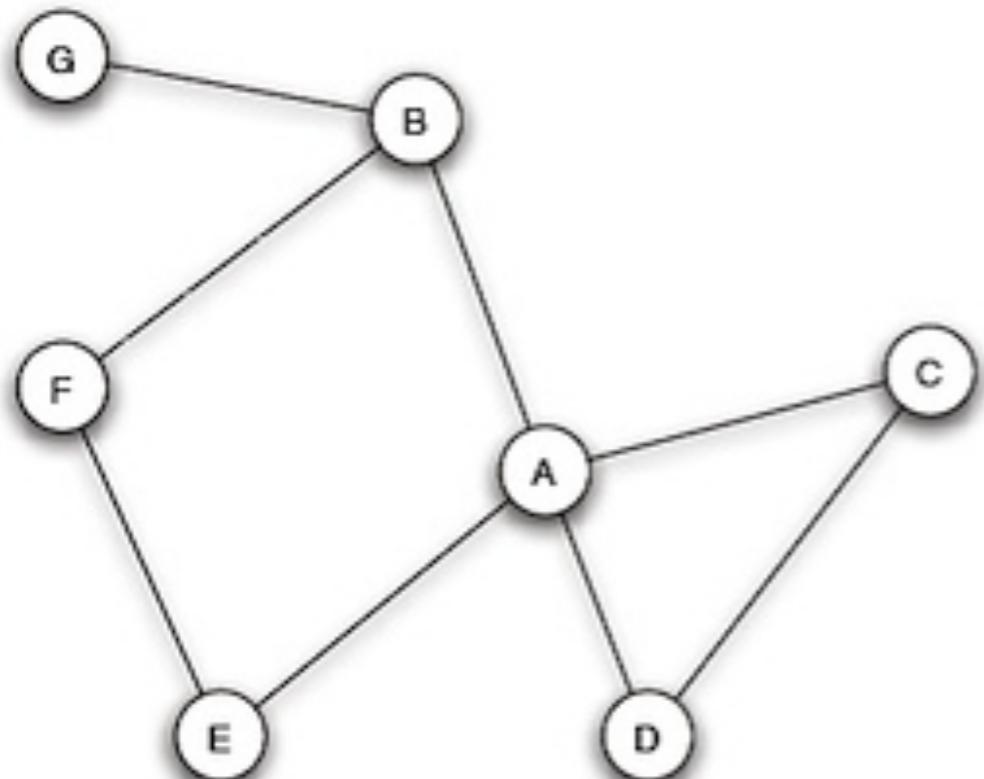
# bridge bridge

A link between otherwise disconnected components

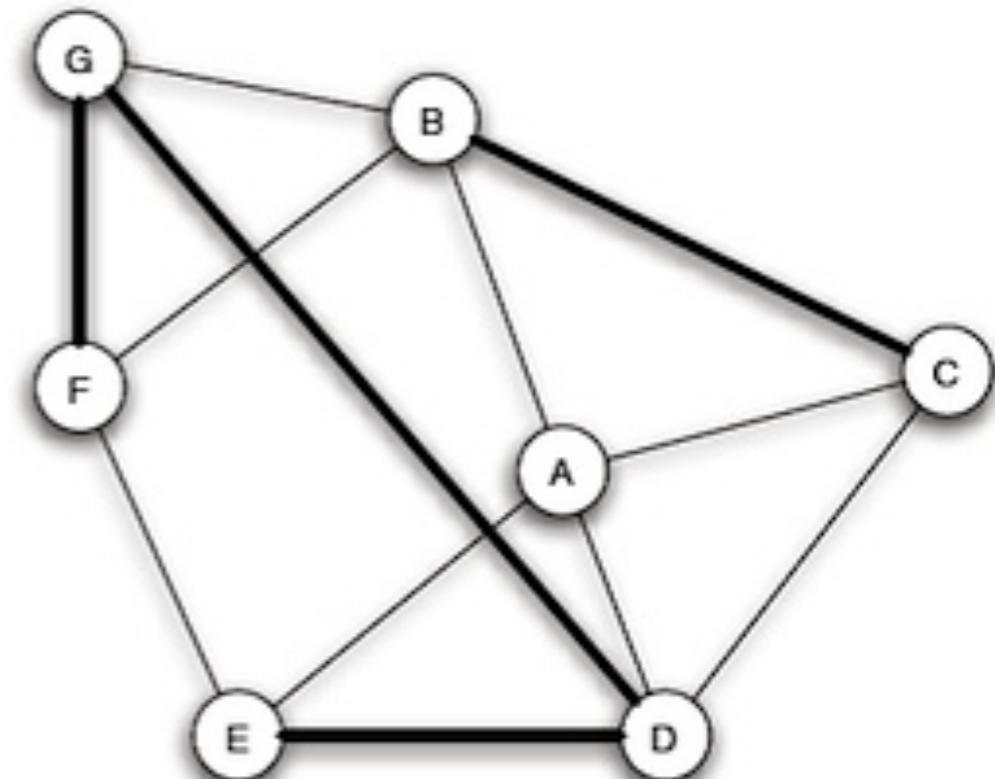


# clustering coefficient

the probability that two randomly selected friends of a node are friends with eachother



(a) Before new edges form.



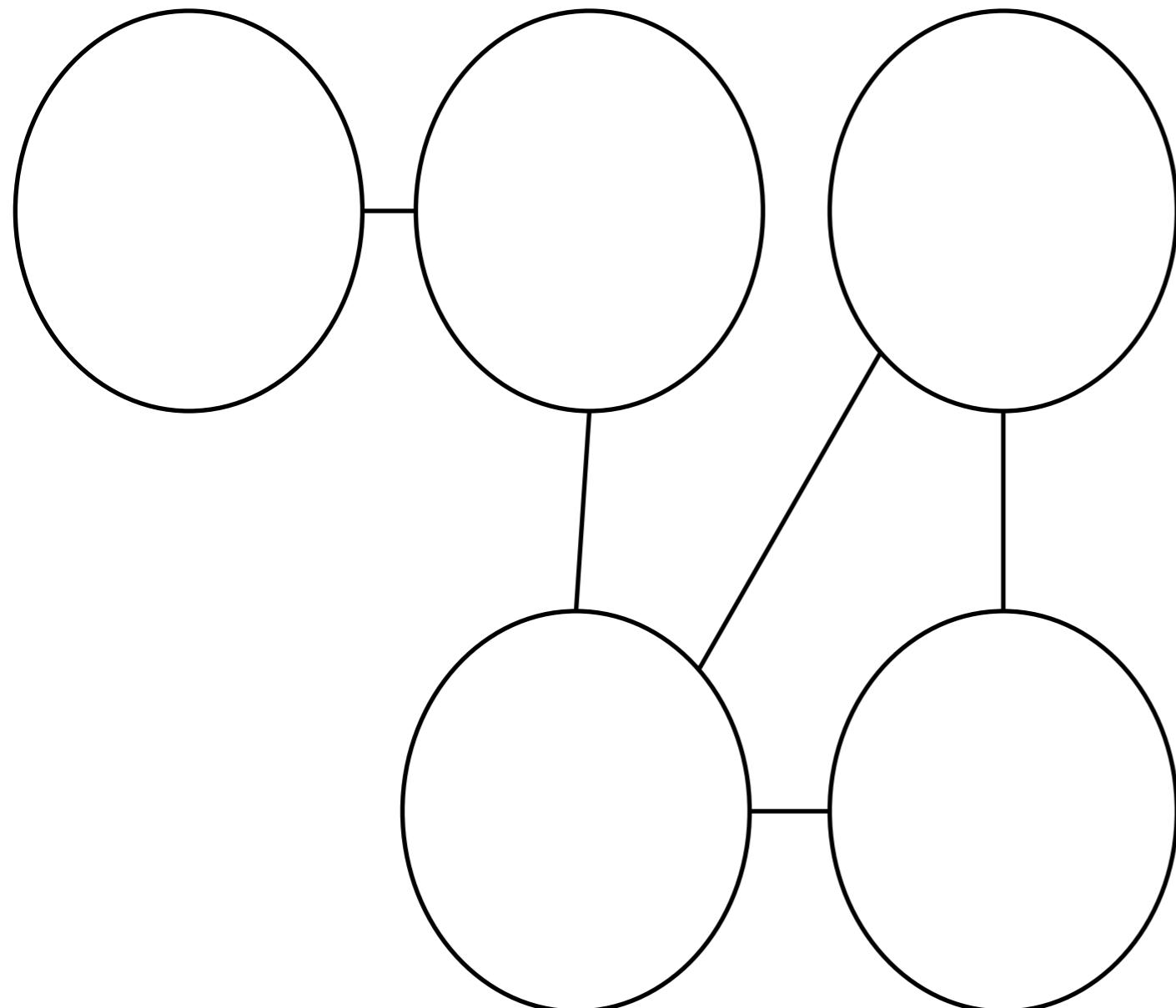
(b) After new edges form.

# centrality



# degree centrality

measure of how connected a node is

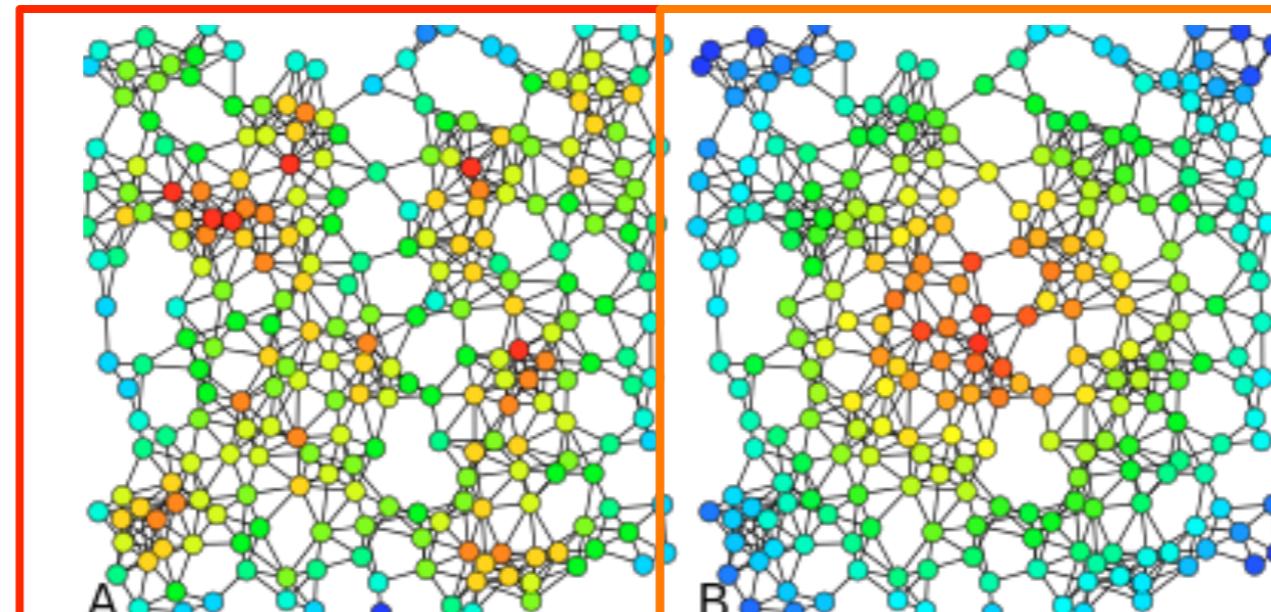


**degree** – no. connections

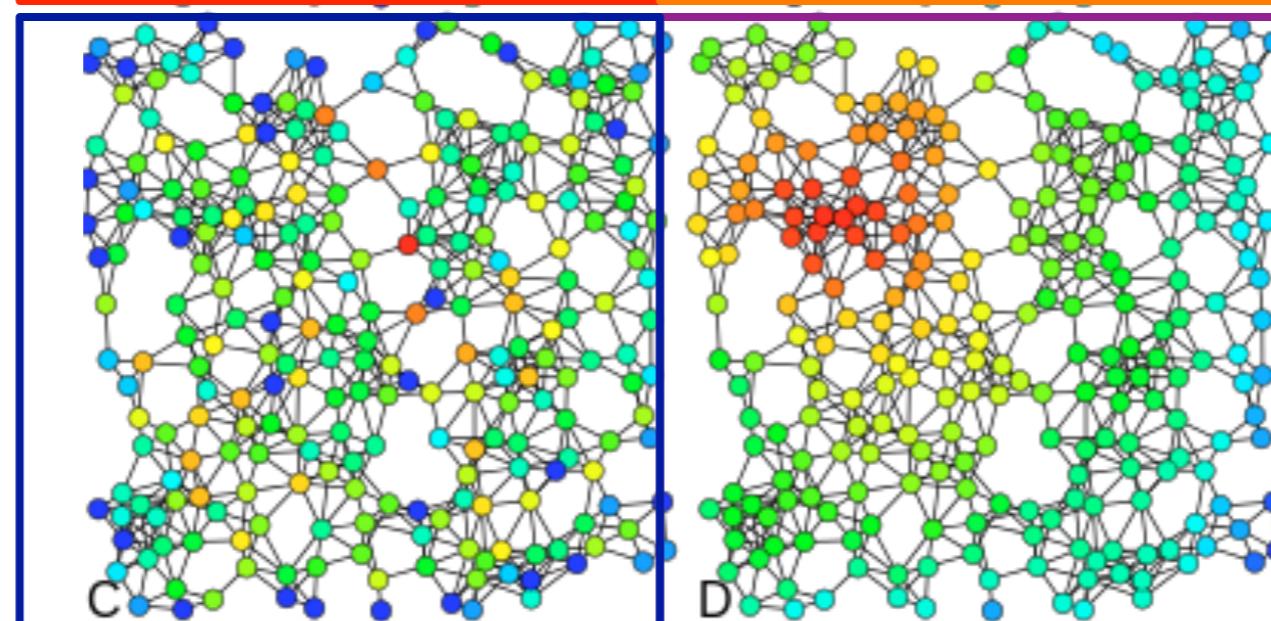
**in-degree** – inbound connections

**out-degree** – outbound connections

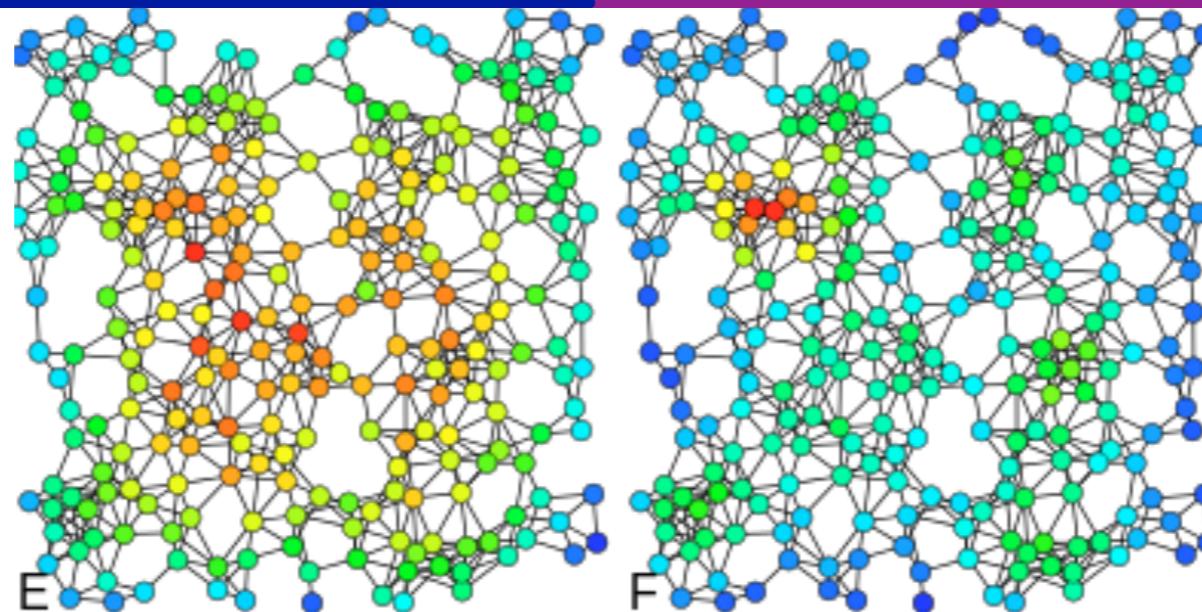
**degree  
centrality**



**betweenness  
centrality**



**closeness  
centrality**



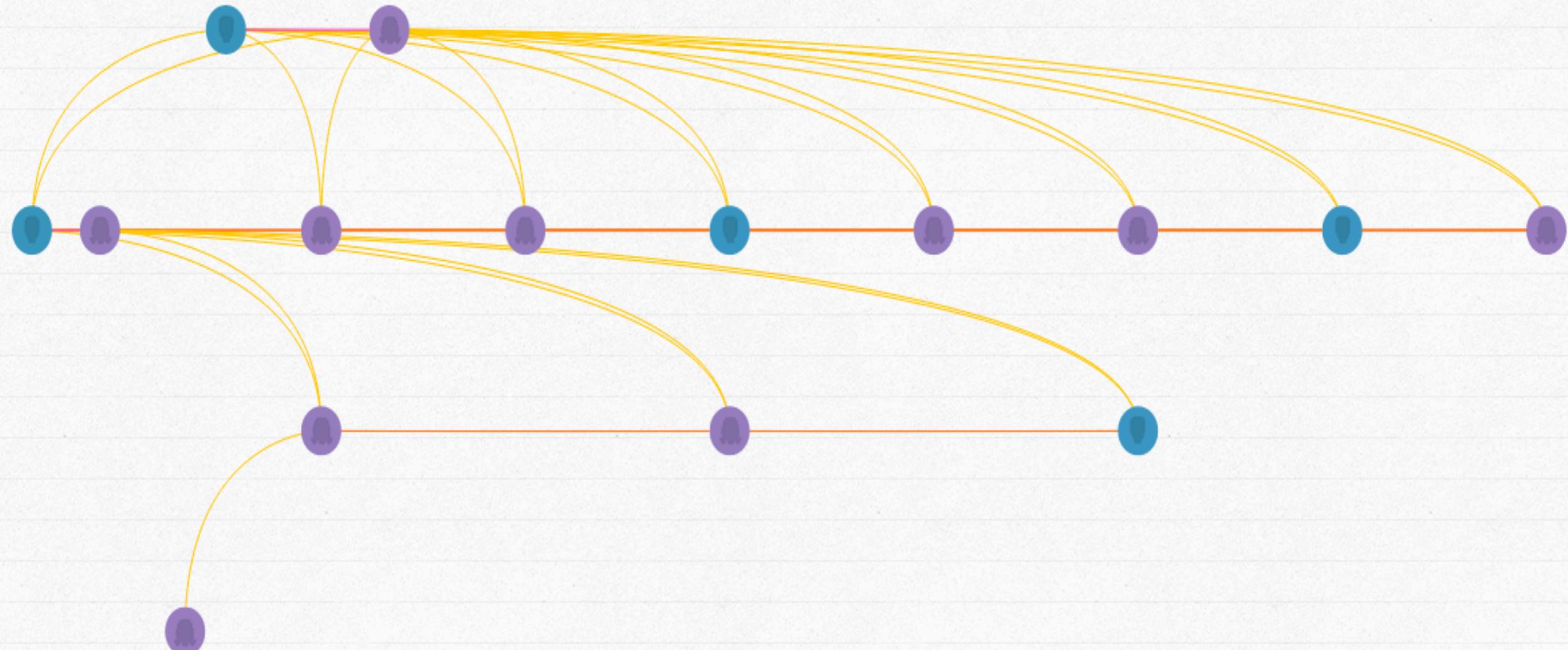
**eigenvector  
centrality**

William Shakespeare  
1564 - 1616



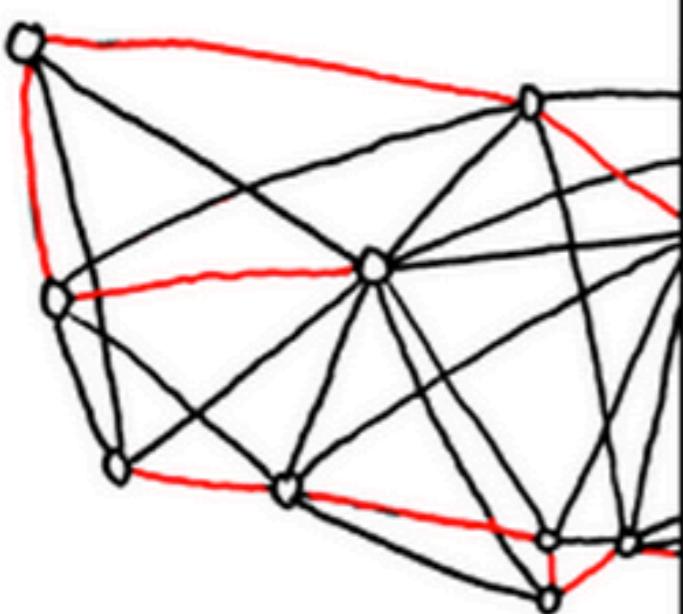
*The family of William Shakespeare*

NETWORK



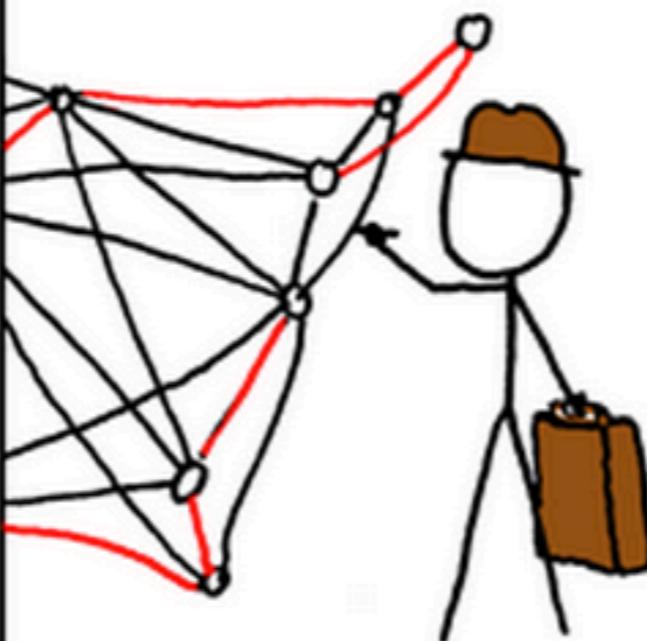
BRUTE-FORCE  
SOLUTION:

$O(n!)$



DYNAMIC  
PROGRAMMING  
ALGORITHMS:

$O(n^2 2^n)$



SELLING ON EBAY:  
 $O(1)$

STILL WORKING  
ON YOUR ROUTE?

SHUT THE  
HELL UP.



# no-no(s)

5  
11 2013

## Networks Demystified 8: When Networks are Inappropriate

• SCOTT WEINGART | □ METHOD | ▷ 2

A few hundred years ago, I promised to talk about when not to use networks, or when networks are used improperly. With [The Historian's Macroscope](#) in the works, I've decided to finally start answering that question, and this Networks Demystified is my first attempt at doing so. If you're new here, this is part of an annoyingly long series (1 network basics, 2 degree, 3 power laws, 4 co-citation analysis, 5 communities and PageRank, 6 this space left intentionally blank, 7 co-citation analysis II). I've issued a lot of vague words of caution without doing a great job of explaining them, so here is the first substantive part of that explanation.

Networks are great. They allow you to do things like understand the role of postal routes in the circulation of knowledge in early modern Europe, or of the spread of the black death in the middle ages, or the diminishing importance of family ties in later Chinese governments. They're versatile, useful, and pretty easy in today's software environment. And they're sexy, to boot. I mean, have you seen this visualization of curved lines connecting U.S. cities? I don't even know what it's supposed to represent, but it sure looks pretty enough to fund!



scott b. weingart

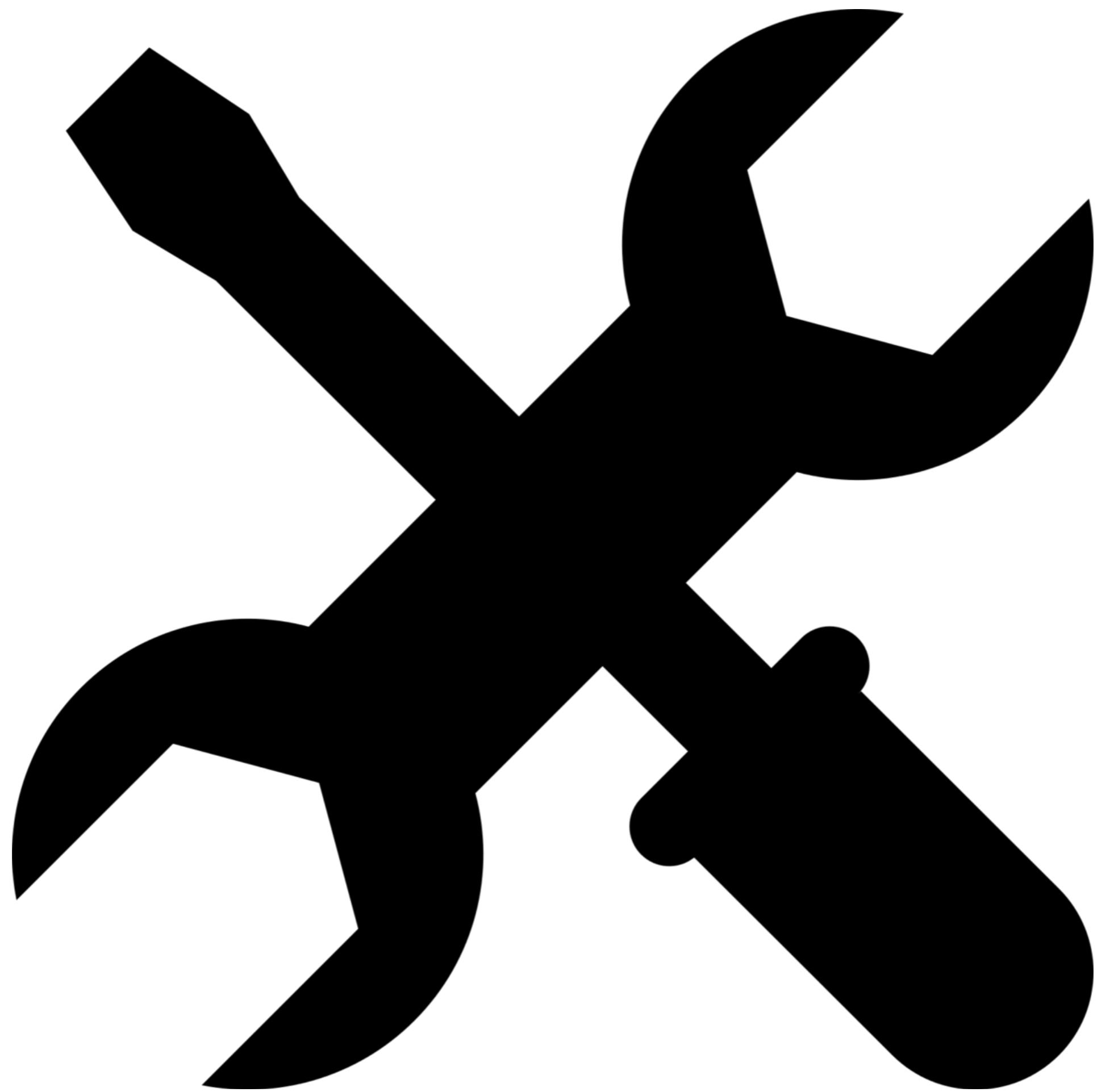
is a juggler, an academic, and a nice guy. He's pretty clueless about a lot of things, and this blog is his attempt to become less so.



alchemy alt-ac appreciability bias big data blogging complexity **data analysis** diffusion digital humanities disciplinarity epistemology experiment gephi GIS history history of science human dynamics



**It's Time For A Break**



# Intro to network data & analysis

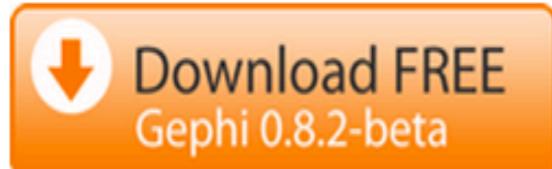
Build  
Visualize  
Measure  
Analyze  
Share

# The Open Graph Viz Platform

Gephi is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs.

Runs on Windows, Linux and Mac OS X. Gephi is open-source and free.

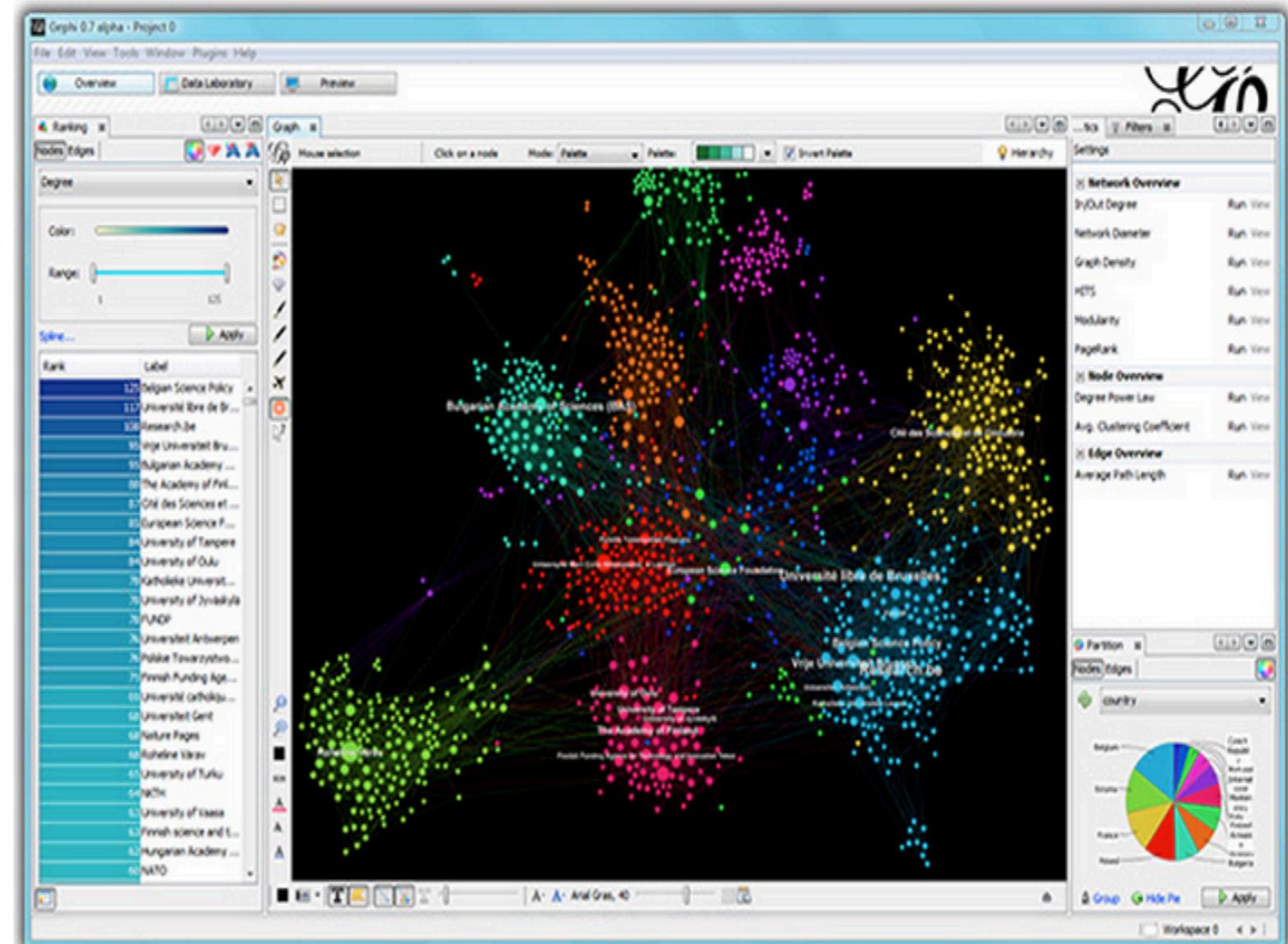
[Learn More on Gephi Platform »](#)



[Release Notes](#) | [System Requirements](#)

► [Features](#)  
► [Quick start](#)

► [Screenshots](#)  
► [Videos](#)





# Data-Driven Documents



# programming language network

A graph of programming languages that consists with their influences, companies, developers, dialects, implementations.

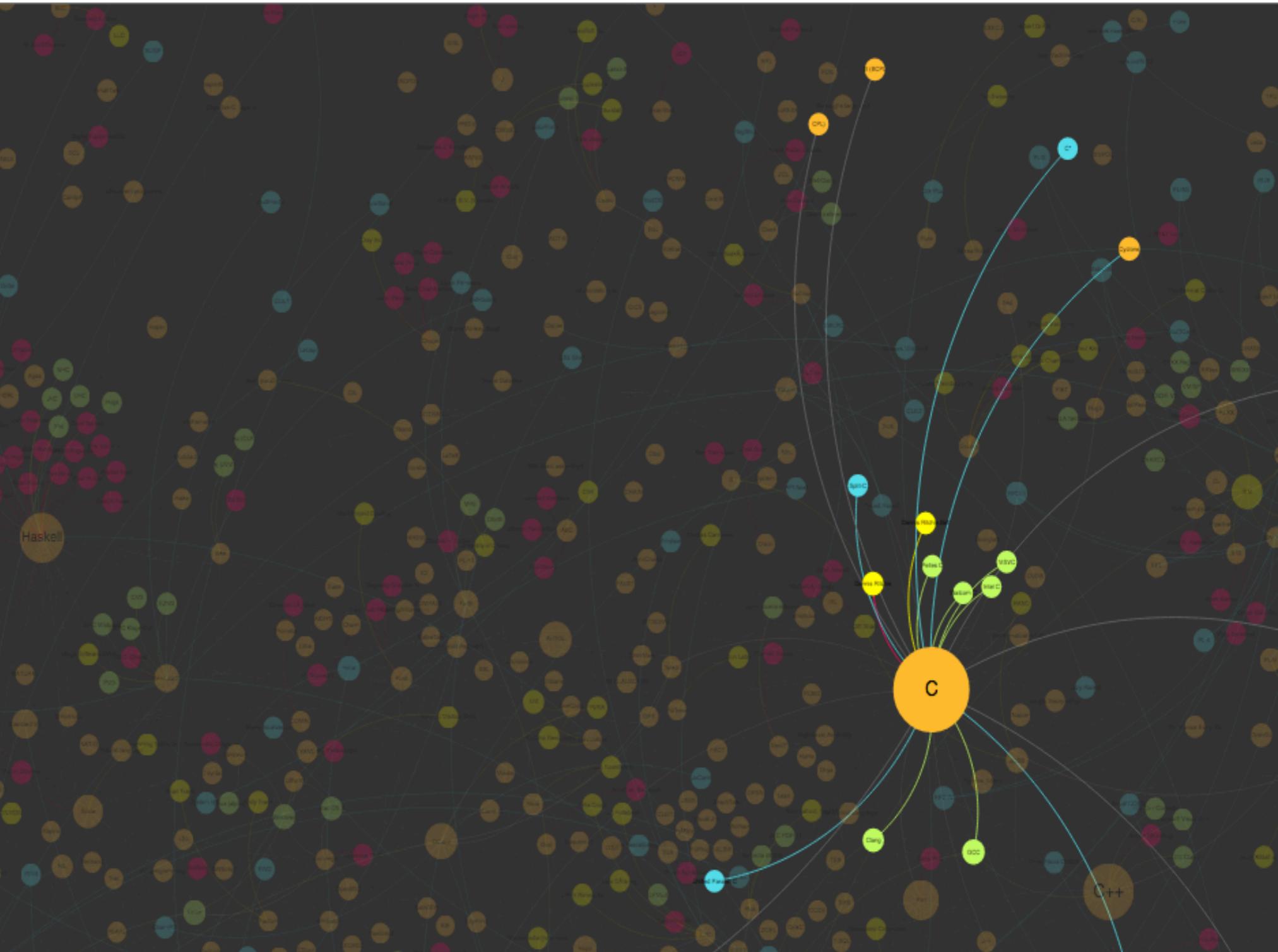
## Nodes

- Programming Language
- Computer Scientist
- Foundation
- Dialect
- Implementation

## Edges

- Influenced by
- Designed by
- Developer
- Dialects
- Major implementations
- Implementation language

source code



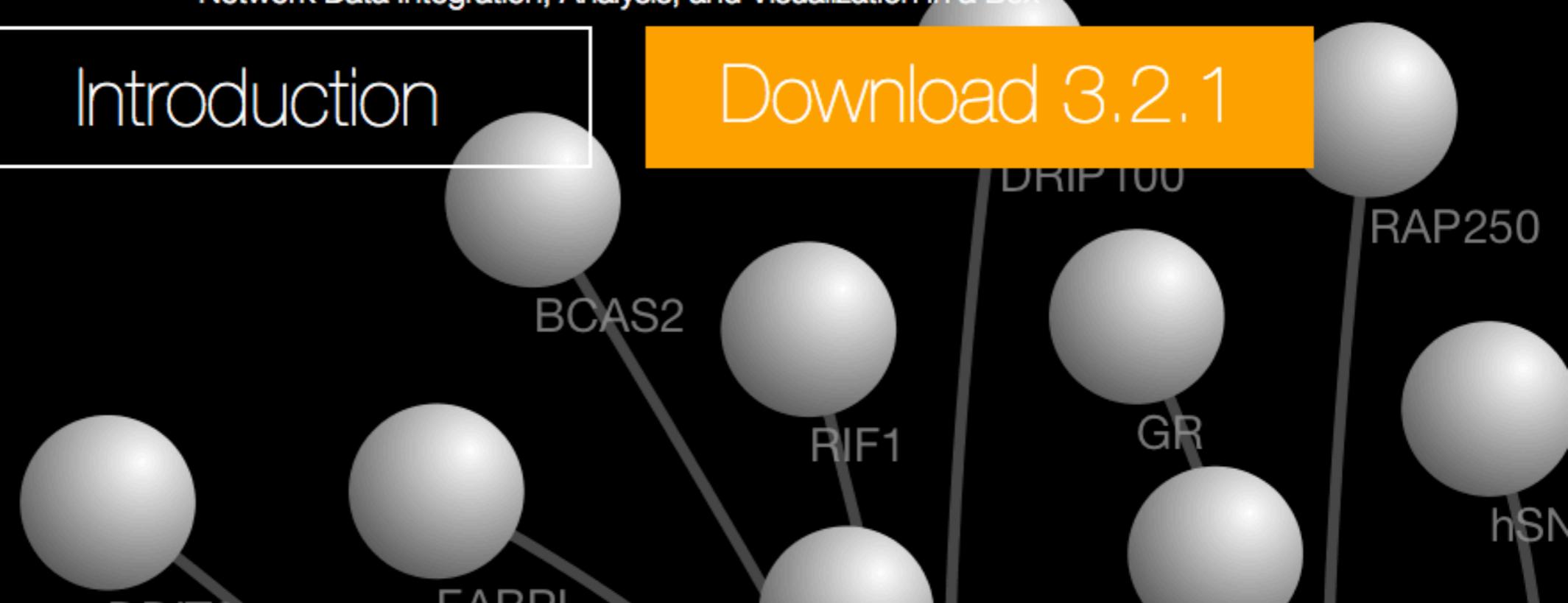


# Cytoscape

Network Data Integration, Analysis, and Visualization in a Box

Introduction

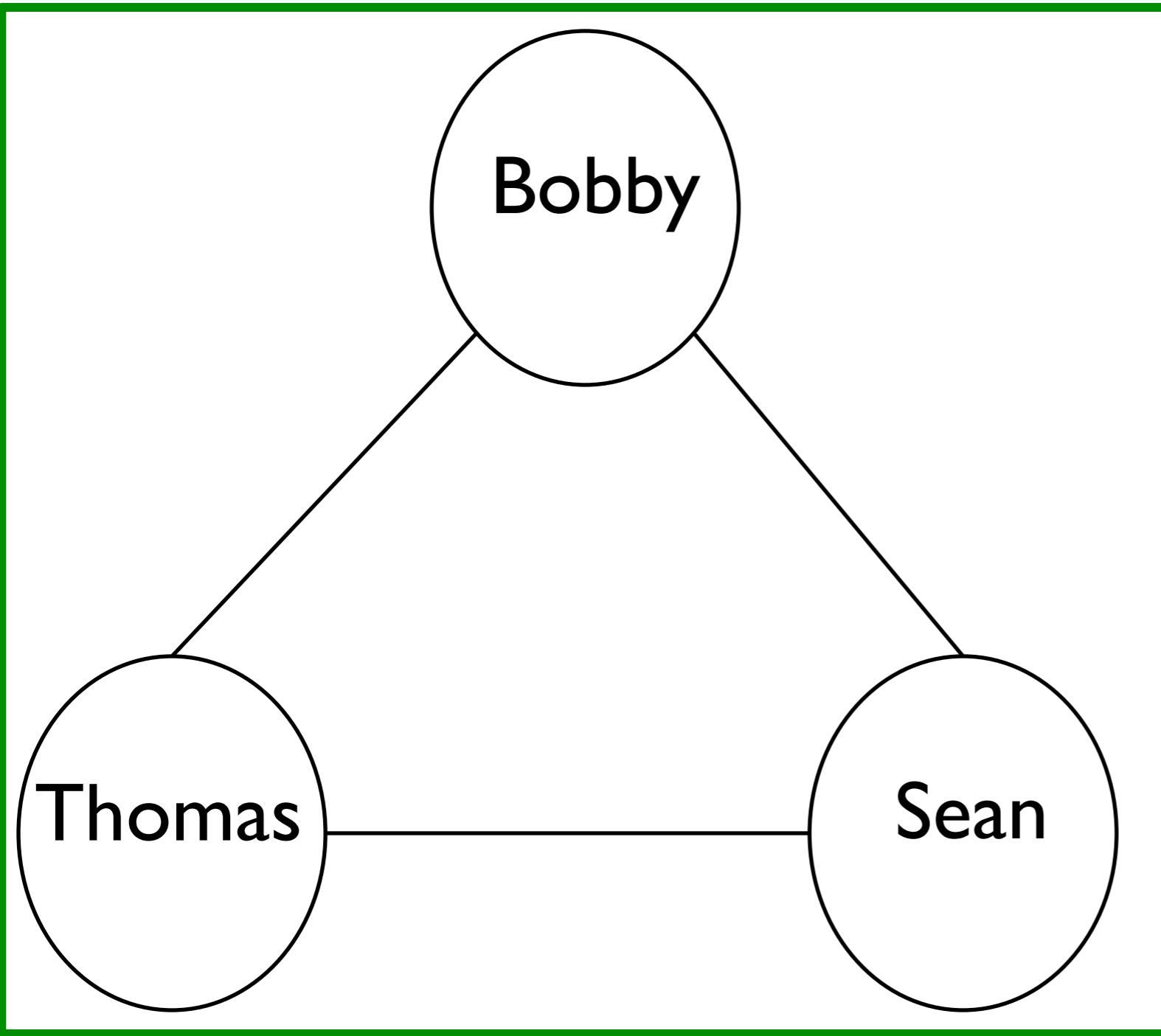
Download 3.2.1



# Data formatting

[wiki.cytoscape.org/Cytoscape User Manual/Creating Networks](http://wiki.cytoscape.org/Cytoscape_User_Manual/Creating_Networks)

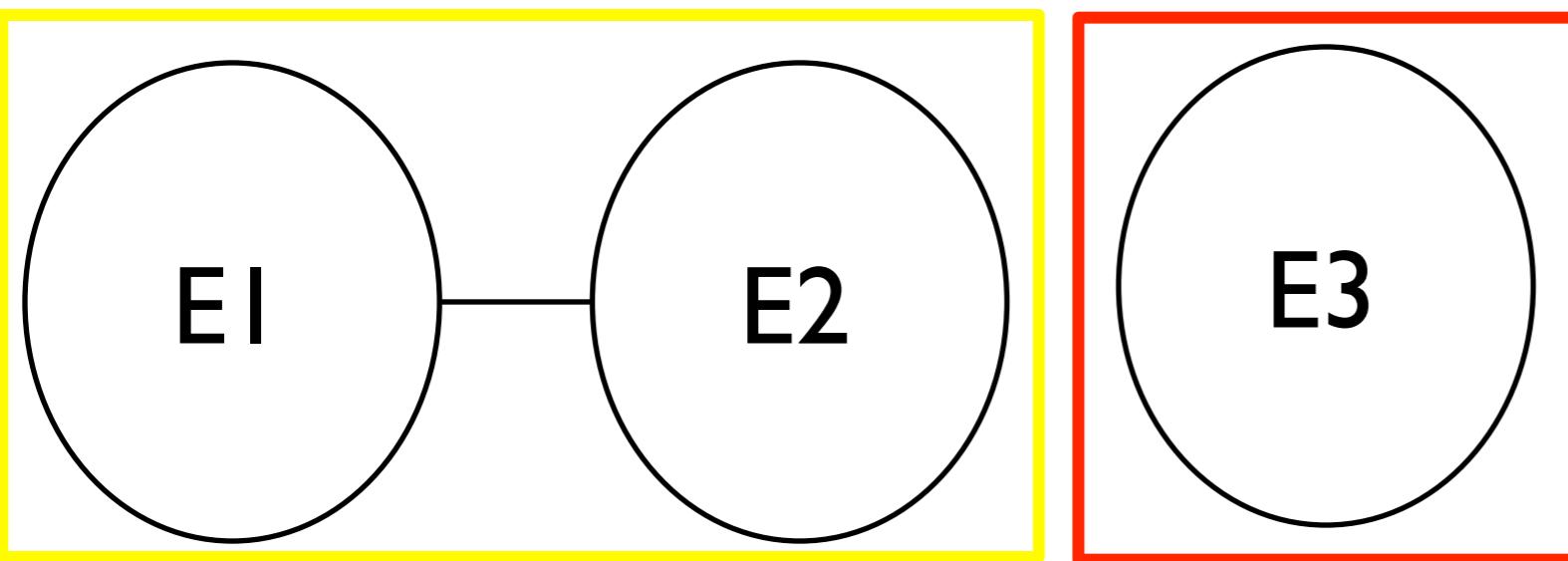
The enemy of my enemy is my friend  
The friend of my friend is my friend  
The friend of my enemy is my enemy  
The enemy of my friend is my enemy



**triad** – group of three dyads

**dyad** – pair of nodes

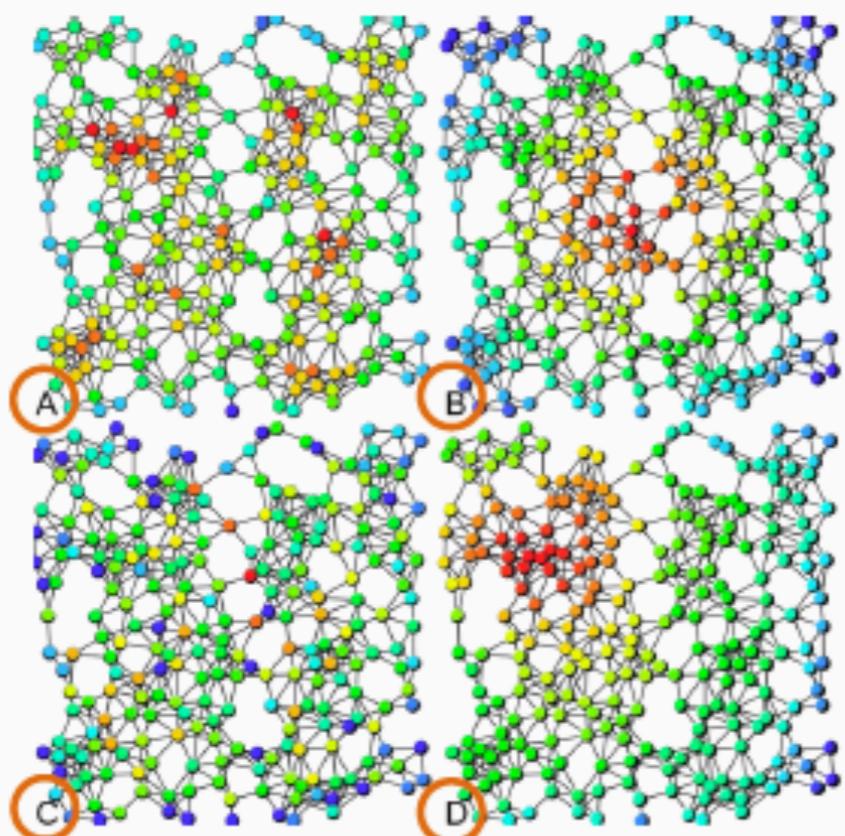
**isolate** – unconnected node



# Download data

[thomaspadilla.org/na2015/data/](http://thomaspadilla.org/na2015/data/)

# Same network, different centralities



Legend:



A: degree centrality = *local connectivity*  
=> nodes having many edges are central

B: closeness centrality = “*geographic*” middle  
=> nodes that are close to all other nodes are central

C: betweenness centrality = *connectivity*  
=> nodes that lay on many shortest paths are central  
(a shortest path is the quickest way to go from one node to another)

D: Eigenvector centrality = *authority*  
=> nodes that are connected to highly ranked nodes are central (recursive approach).

Ref: <http://en.wikipedia.org/wiki/File:Centrality.svg>