APPLIED DATA SCIENCE

fall 2017

Session 10: Introduction to Network Analysis. Node centrality

Instructor: Prof. Stanislav Sobolevsky
Course Assistants: Tushar Ahuja, Maxim Temnogorod

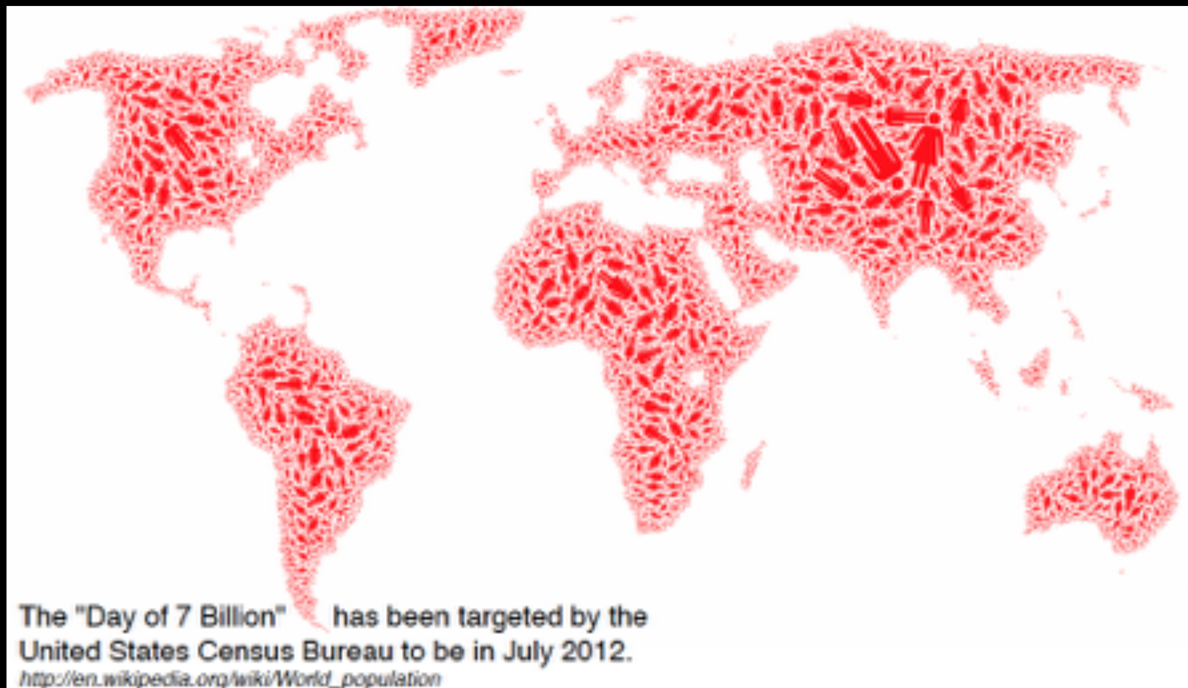
Complexity

COMPLEX

[adj., v. kuh m-pleks, kom-pleks; n. kom-pleks]
–adjective
1.
composed of many interconnected parts; compound; composite: a complex highway system.
2.
characterized by a very complicated or involved arrangement of parts, units, etc.: complex machinery.
3.
so complicated or intricate as to be hard to understand or deal with: a complex problem.
Source: dictionary.com

Complexity, a scientific theory which asserts that some systems display behavioral phenomena that are completely inexplicable by any conventional analysis of the systems' constituent parts. These phenomena, commonly referred to as emergent behaviour, seem to occur in many complex systems involving living organisms, such as a stock market or the human brain.
Source: John L. Casti, Encyclopædia Britannica

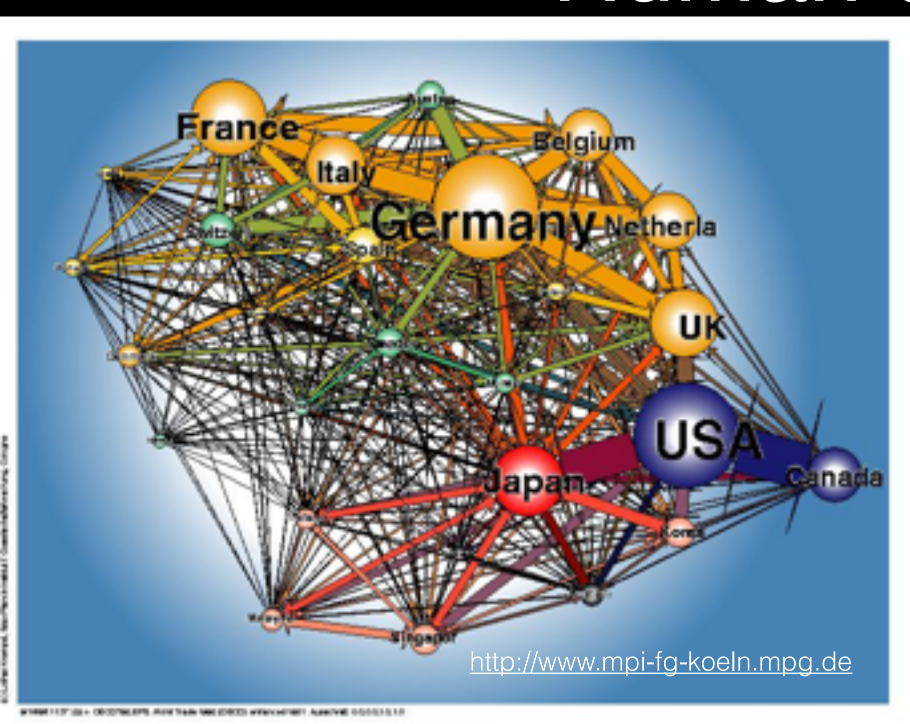
Composition from: http://barabasi.com/networksciencebook

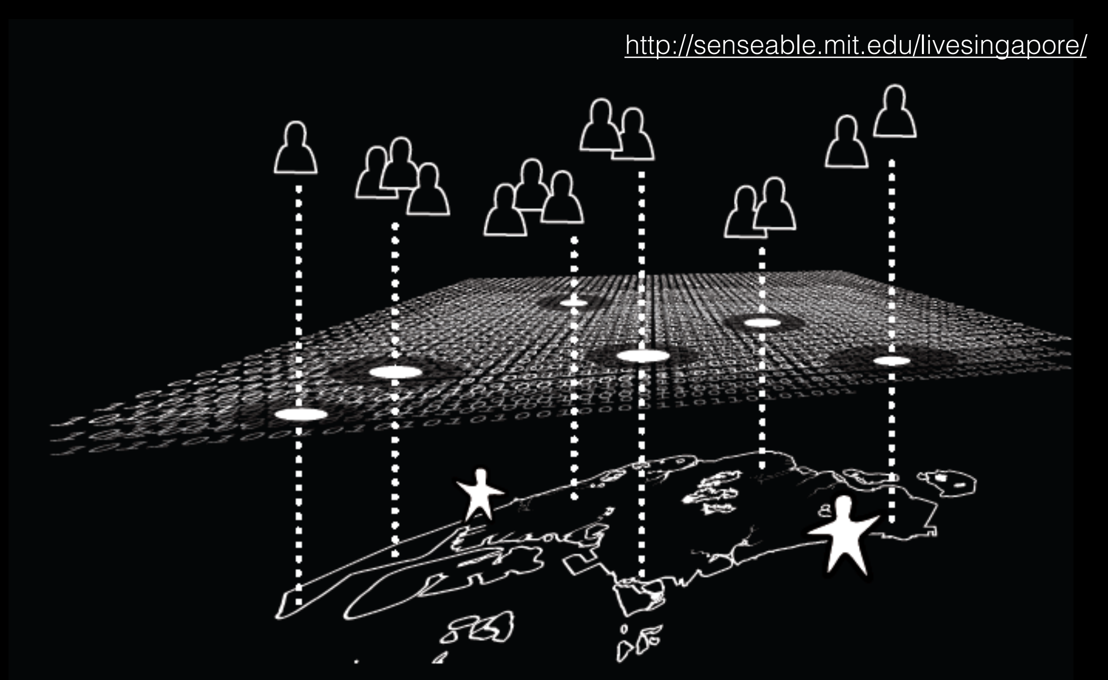


Human body *Human society*

Complexity

Human economy *Human cities*





Complex network

Behind each complex system there is a **network**

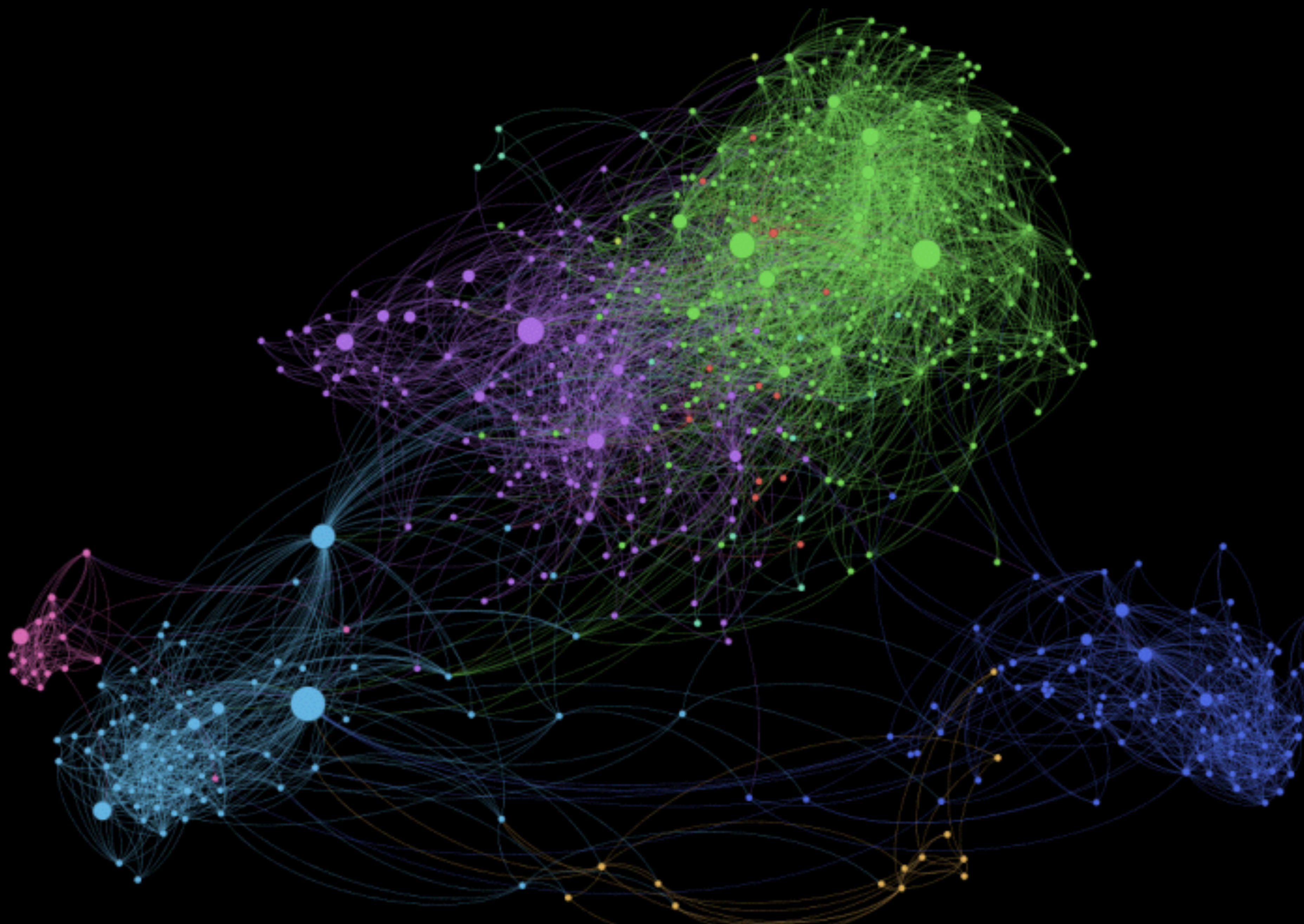
A.-L. Barabasi **Network Science : Introduction 2012**

Network examples - human society

http://iconof.com/blog/visualizing-friendships-on-facebook/

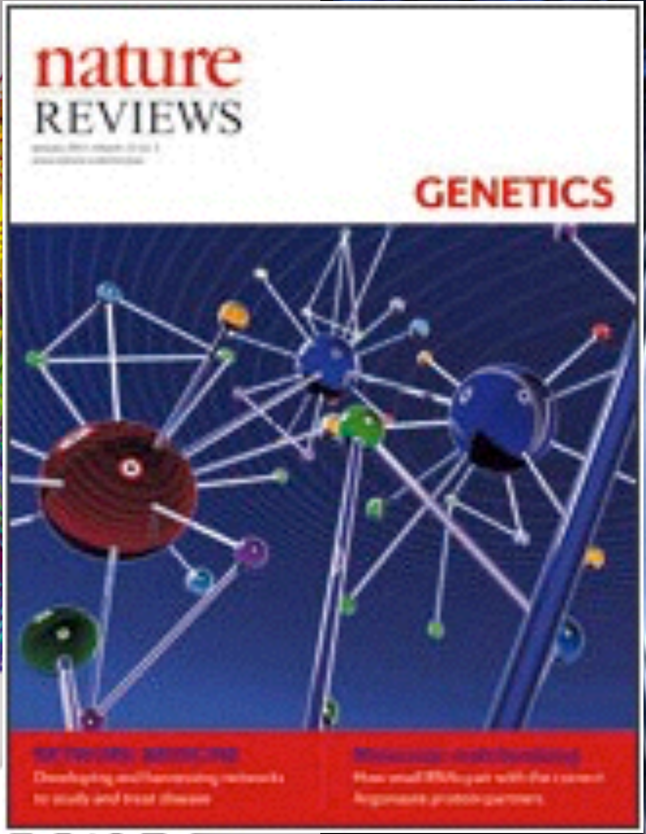
7.3B people, >7.3B cell phones, 1.5B - Facebook, 400M - LinkedIn

Network examples - human society

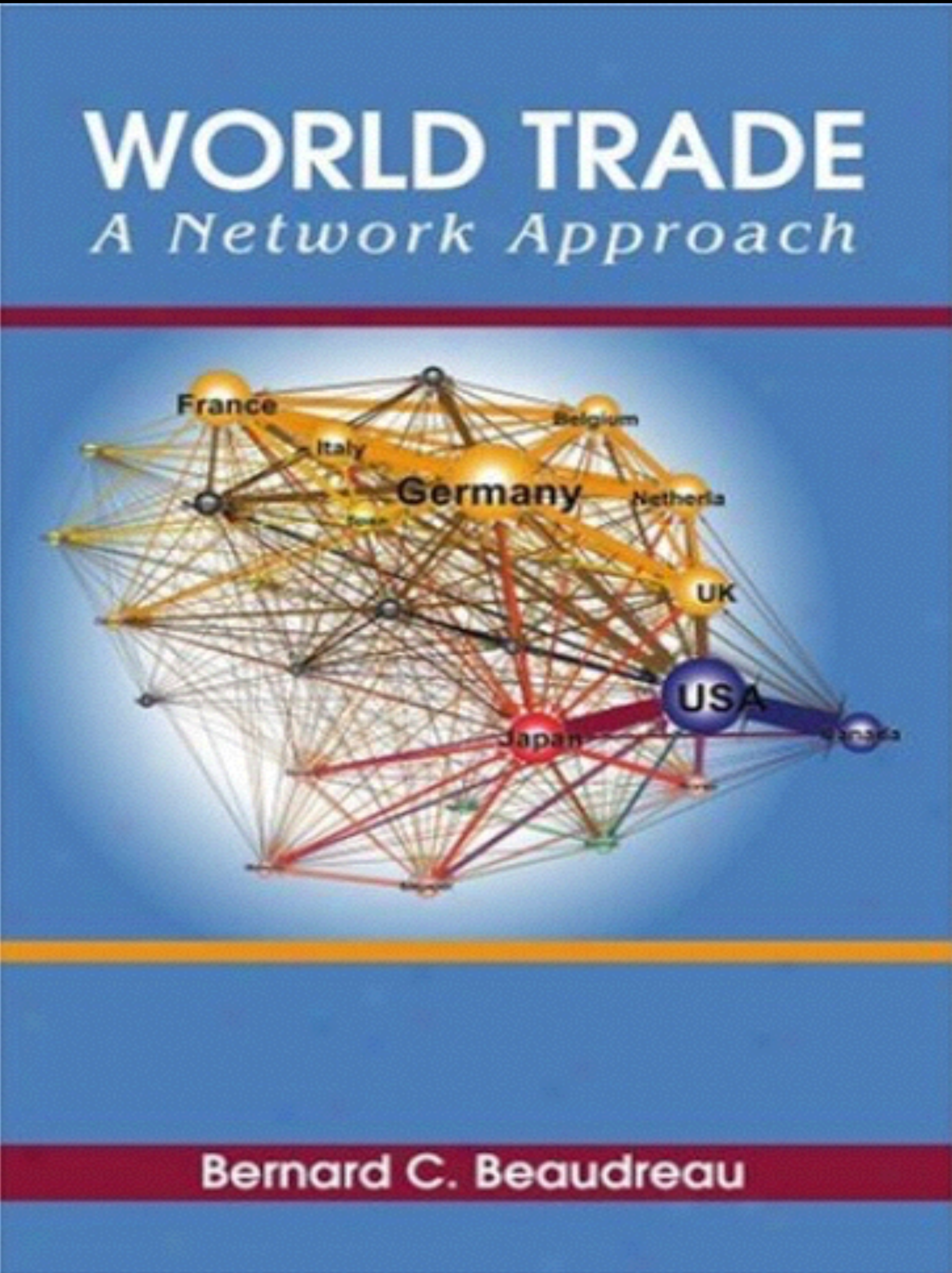


http://kimoquaintance.com/2011/08/22/what-can-we-learn-about-somalis-from-their-facebook-networks/

Network examples - human body

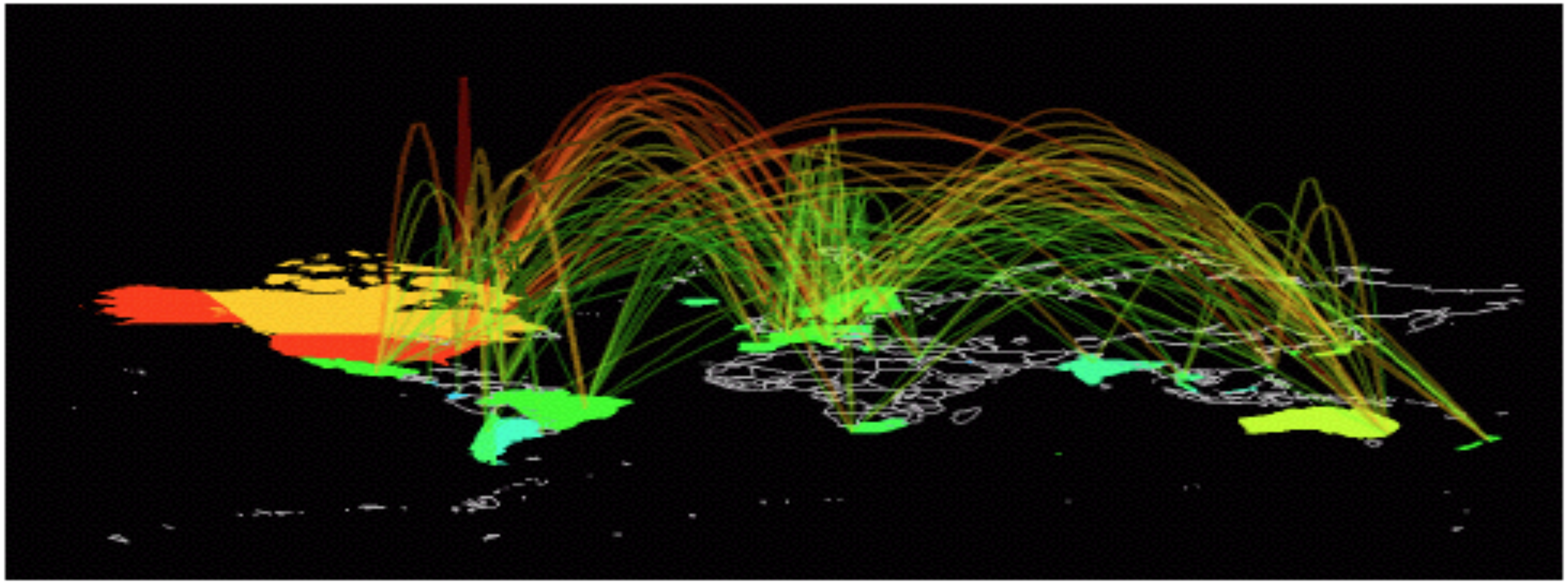


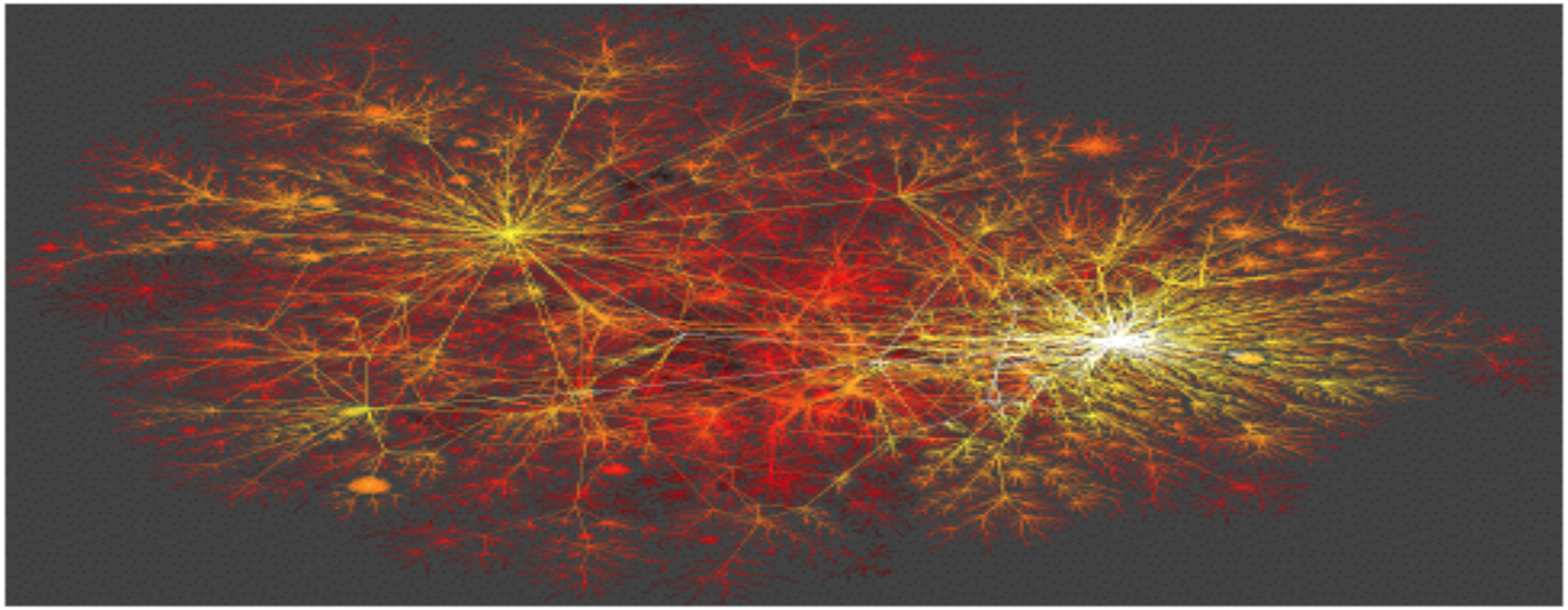
Network examples - human economy

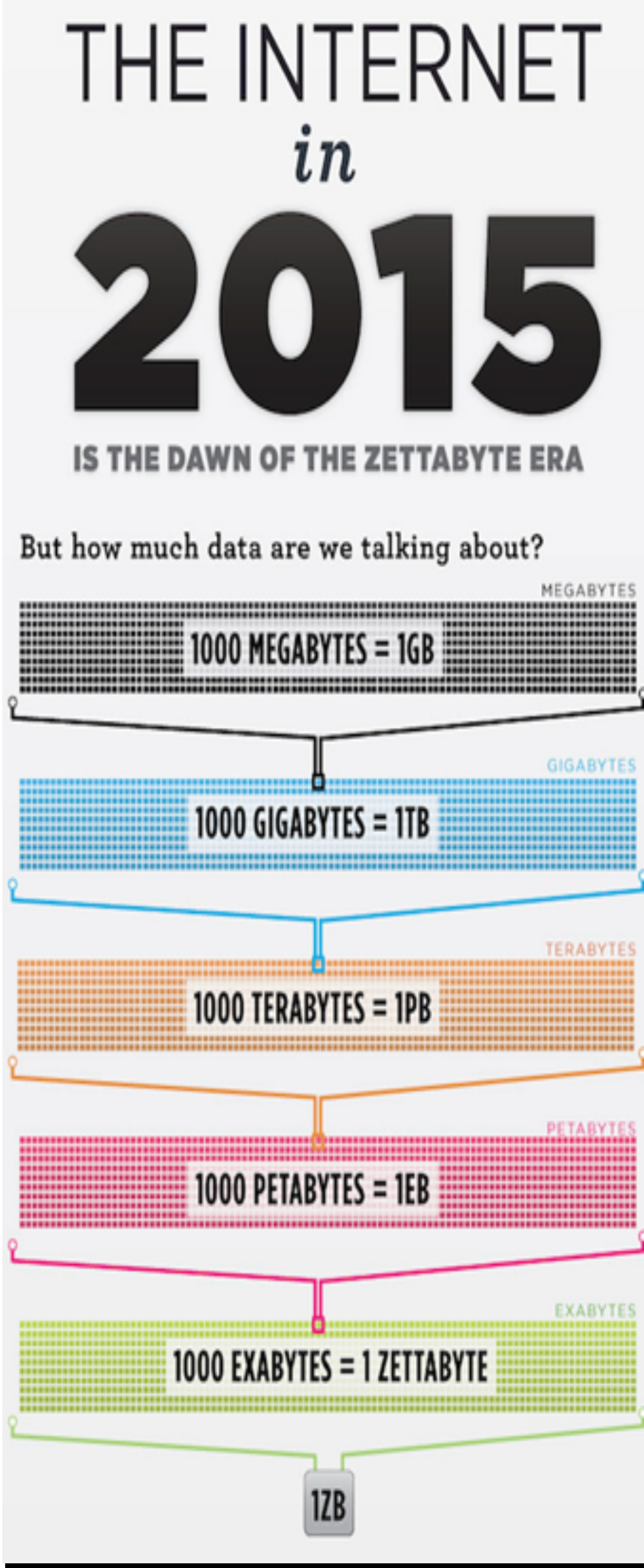


www.wto.org

Network examples - internet







http://barabasi.com/networksciencebook

Network Science: Introduction January 10, 2011

http://www.theguardian.com/technology/blog/2011/jun/29/zettabyte-data-internet-cisco

Economic impact

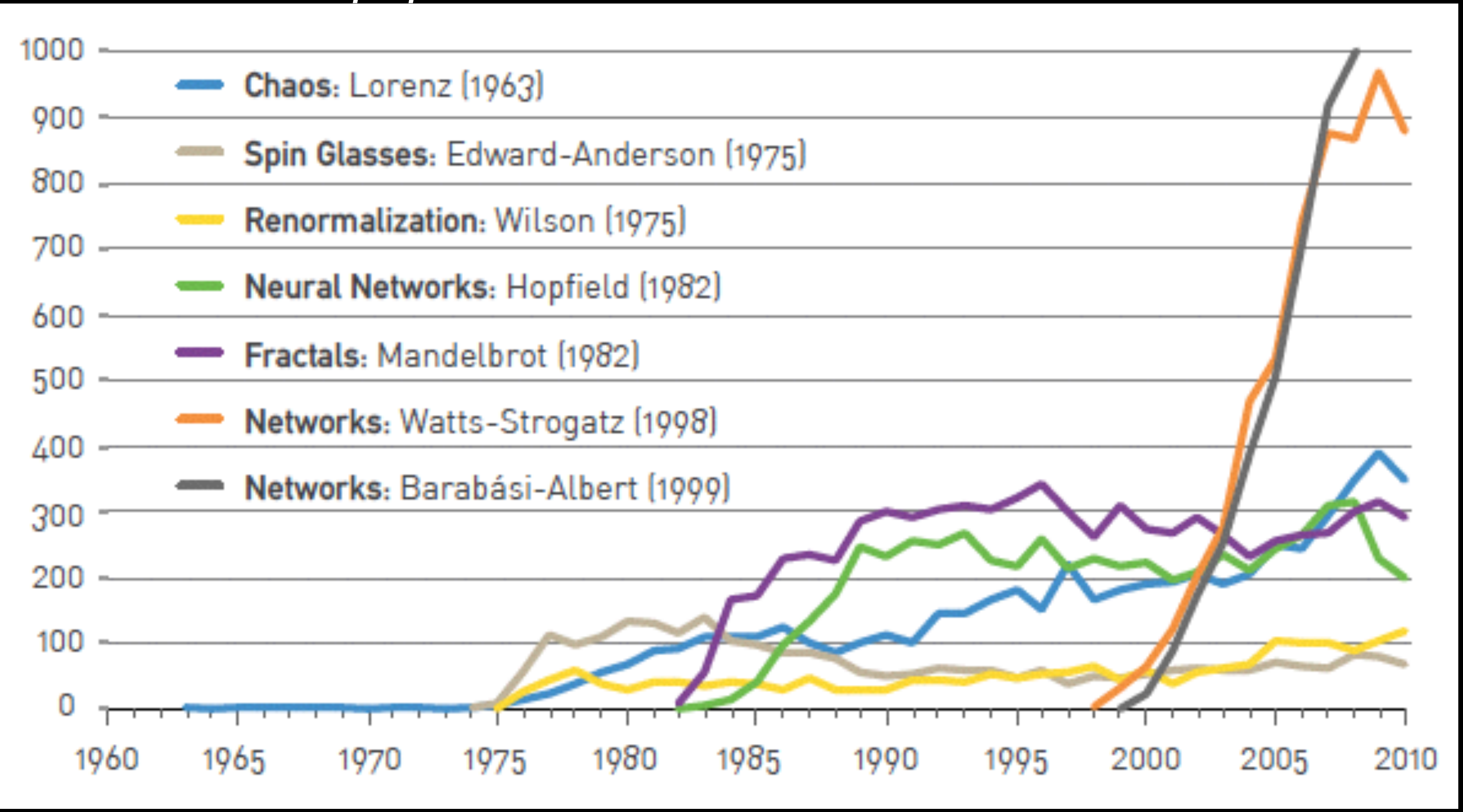
Google - 500B
Facebook - 300B
Cisco - 150B
Yahoo - 50B
LinkedIn -30B
UBER - 40-50B

Pharmacology
Healthcare: Genetics
Finance
Transportation
Urban planning

Scientific impact

Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of ‘small-world’networks." *nature* 393.6684 (1998): 440-442.
27453 citations
second most cited paper in nature from 1998

Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *science* 286.5439 (1999): 509-512.
23759 citations
first most cited paper in science from 1999

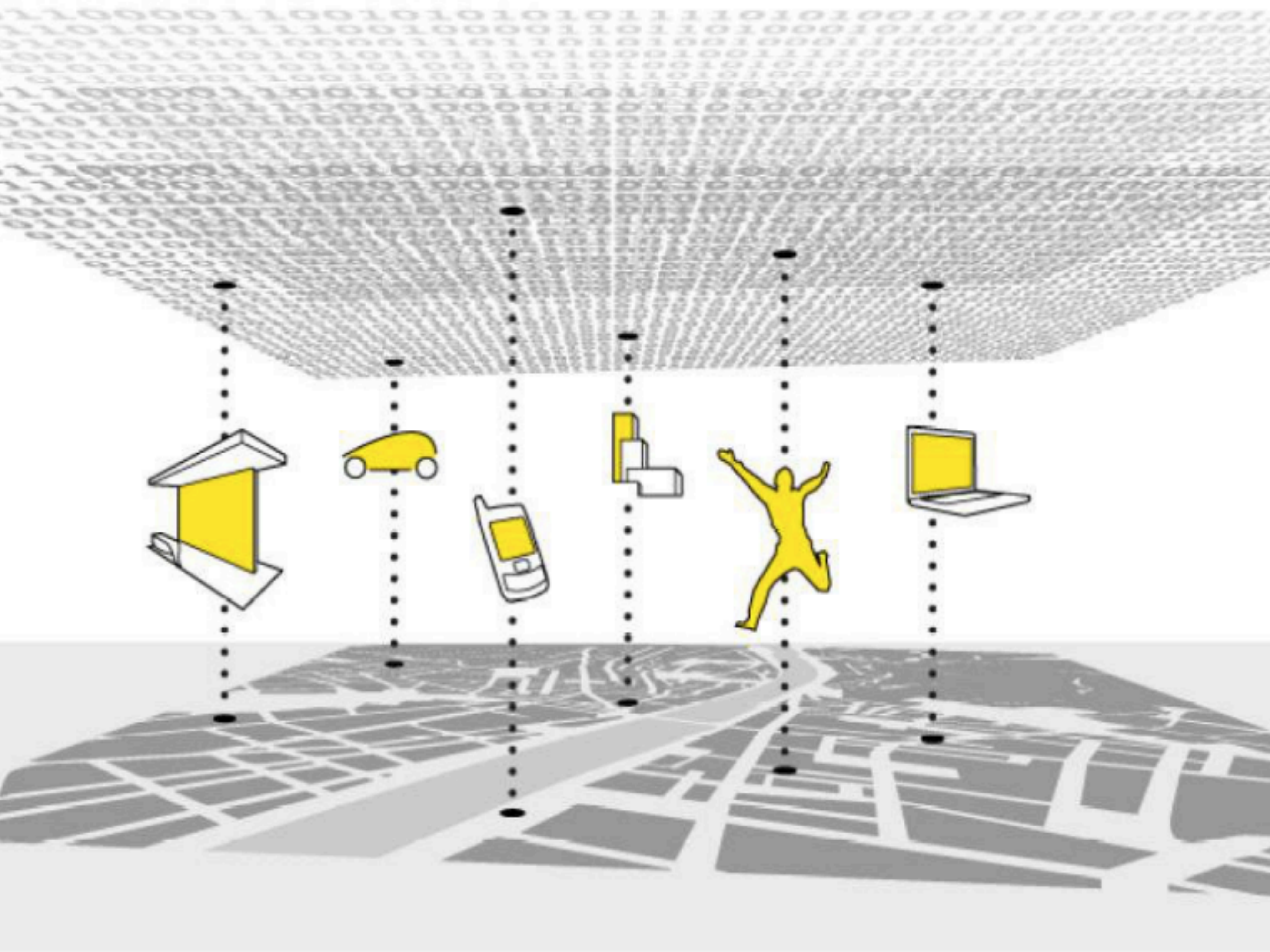


http://barabasi.com/networksciencebook

Drivers of network theory

Graph theory (XVIII century)
Social network research: early XX century
Communication networks (1960s), Internet
Biological networks: 1970s

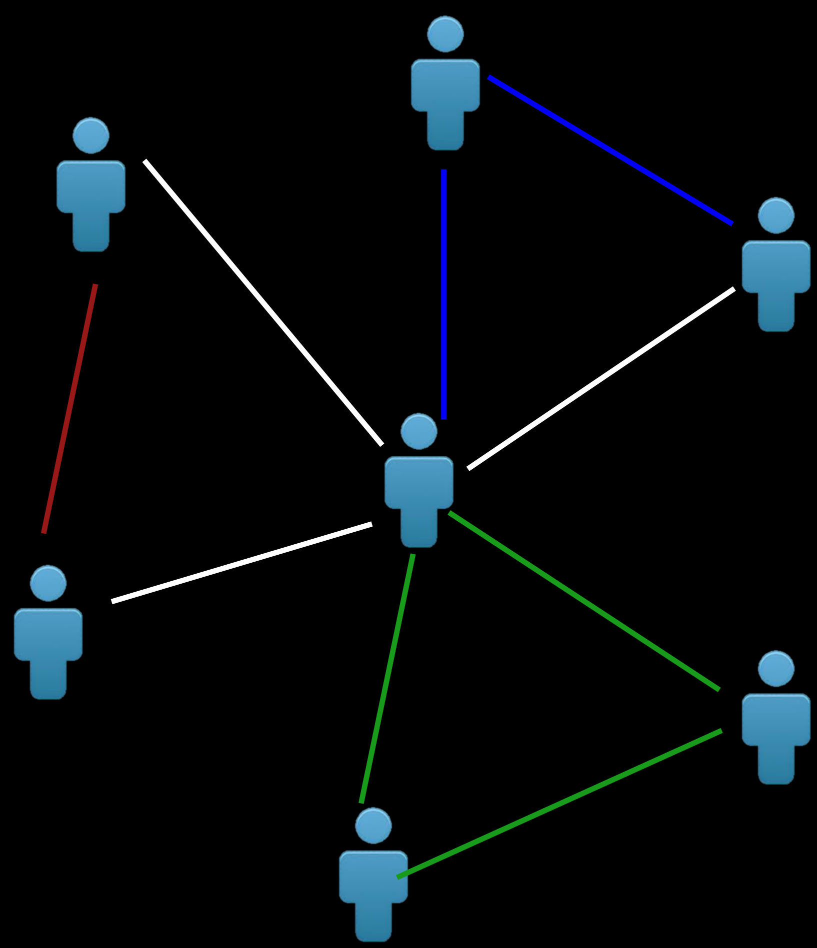
Network science: 1990s, Barabasi, Strogatz, Newman…



Urban datasets

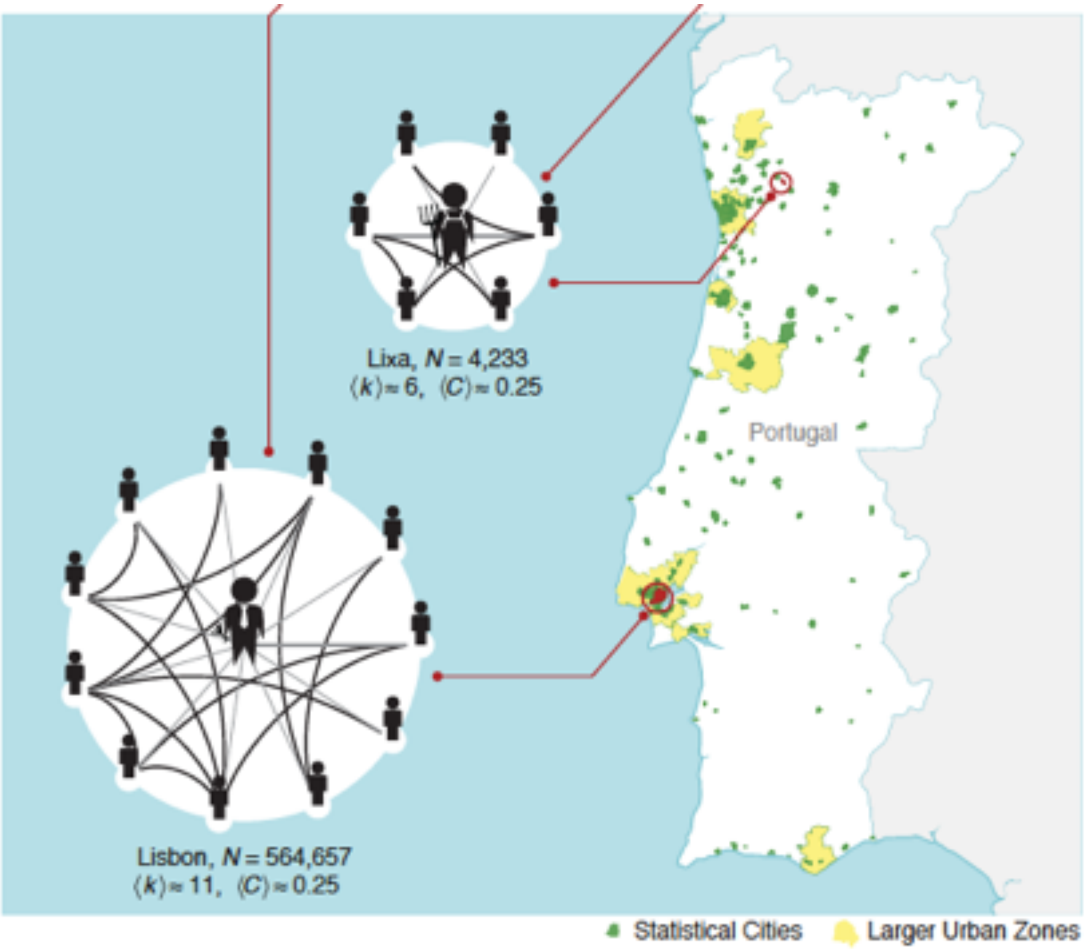
- Public transportation check-ins/outs
- Taxi ridership
- 311 complains; 911 calls
- Pedestrian and traffic counts
- Energy usage
- Waste collection
- Environmental: weather, air quality, noise levels
- Real-estate sales
- Social media: Twitter, Instagram, Foursquare, Flickr, Yelp
- Airbnb
- WiFi usage
- Cell phone data
- Credit card transactions

Individual relations



- Friendship
- Romantic
- Family ties
- Professional collaboration

Individual human interactions

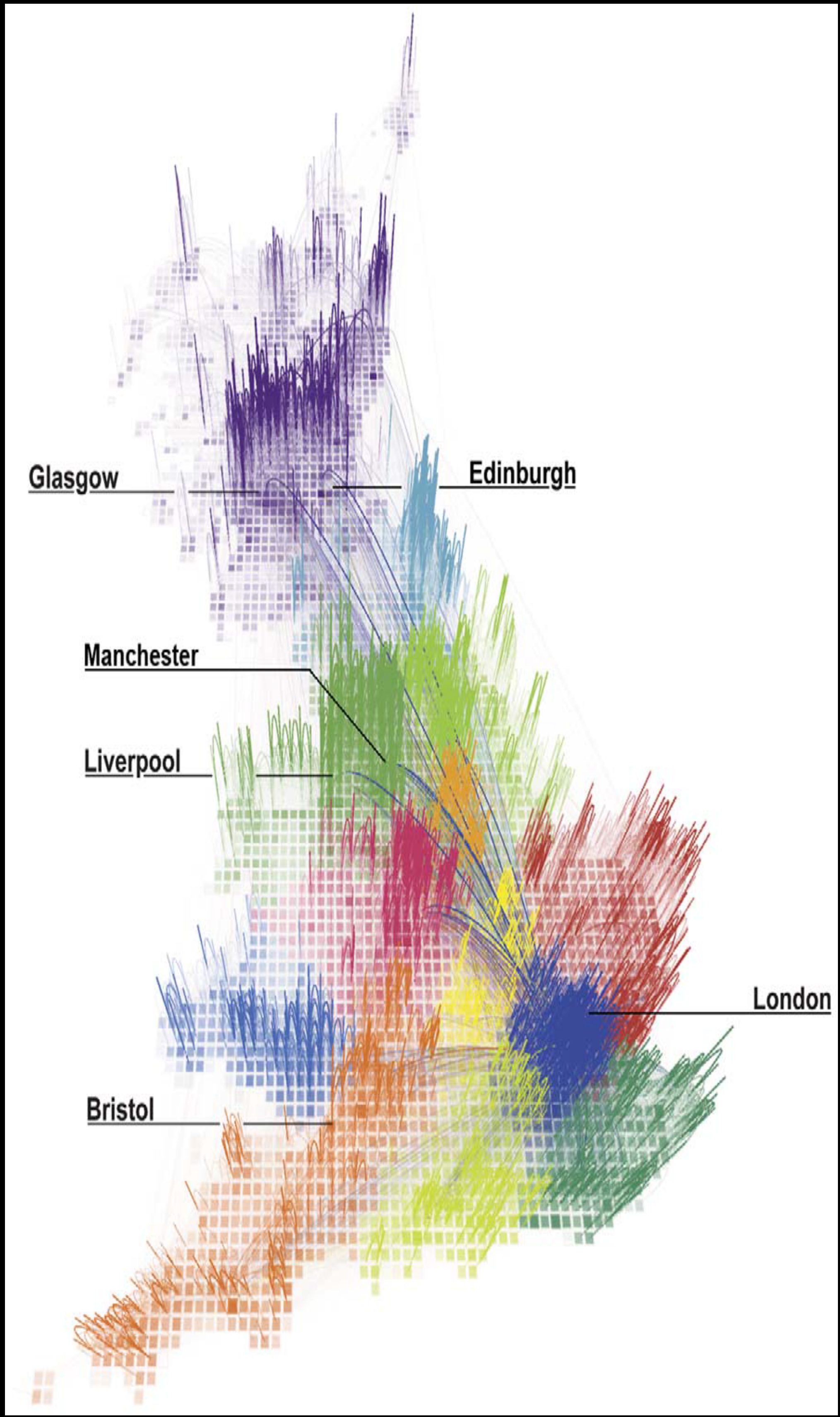


Schläpfer, M., Bettencourt, L. M., Grauwin, S., Raschke, M., Claxton, R., Smoreda, Z., ... & Ratti, C. (2014). The scaling of human interactions with city size. *Journal of The Royal Society Interface*, *11*(98), 20130789.

http://barabasi.com/networksciencebook

Aggregated human interactions

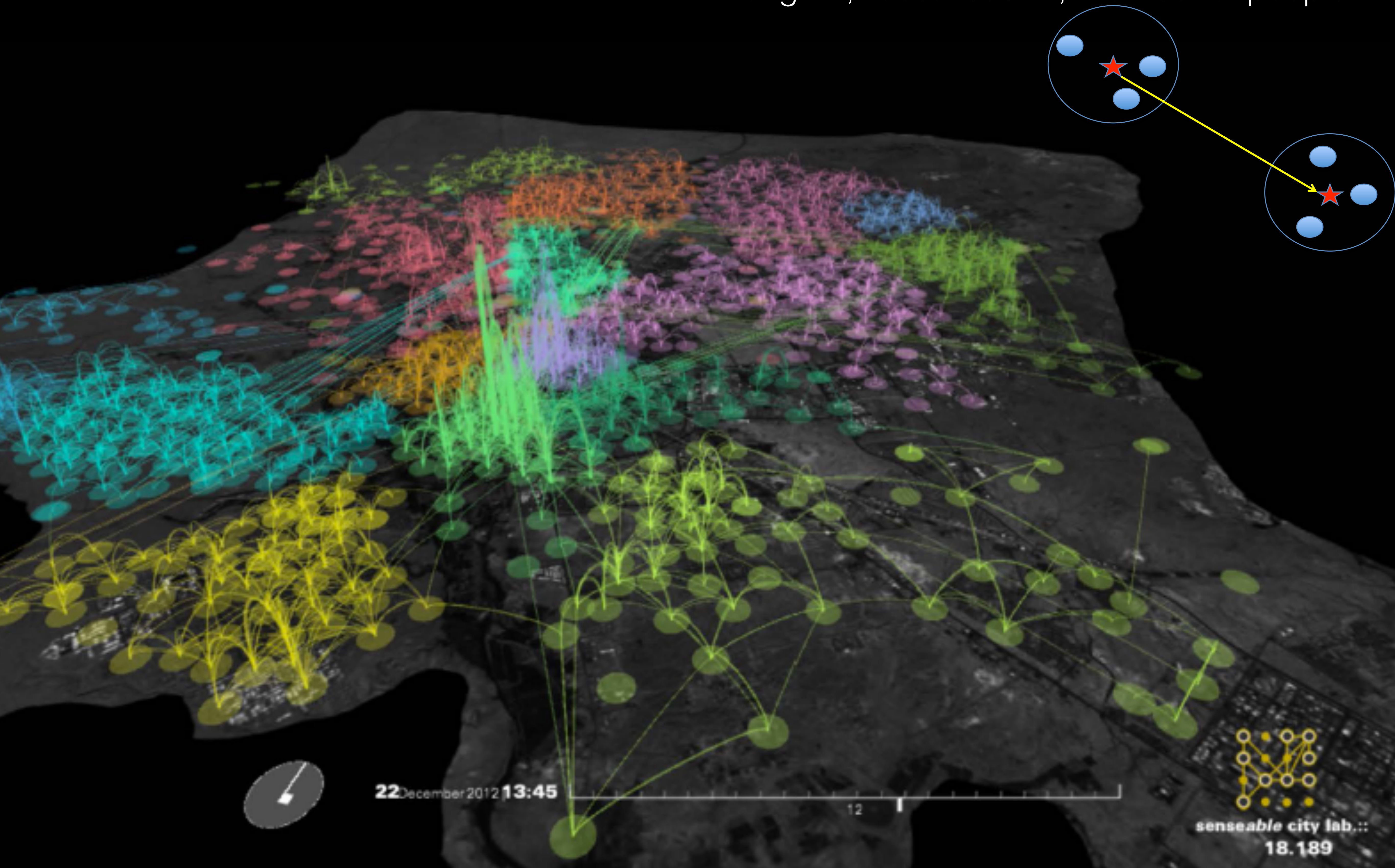
Communication fluxes:
<origin>,<destination>,<number/duration of calls>



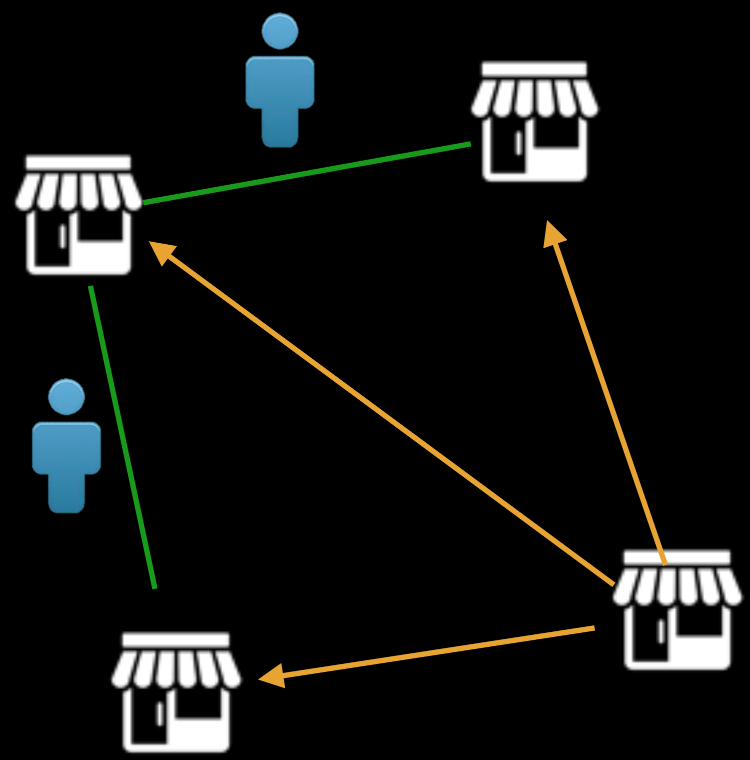
Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., ... & Strogatz, S. H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PloS one*, *5*(12), e14248.

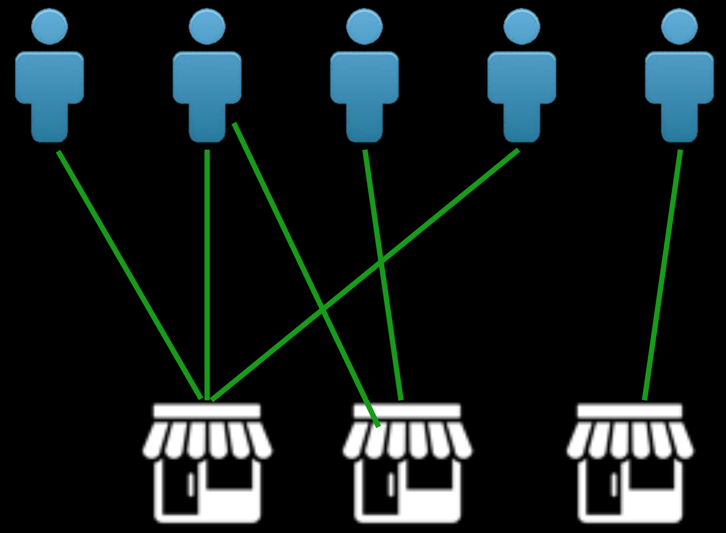
Human mobility networks

Mobility fluxes:
<origin>,<destination>,<number of people>



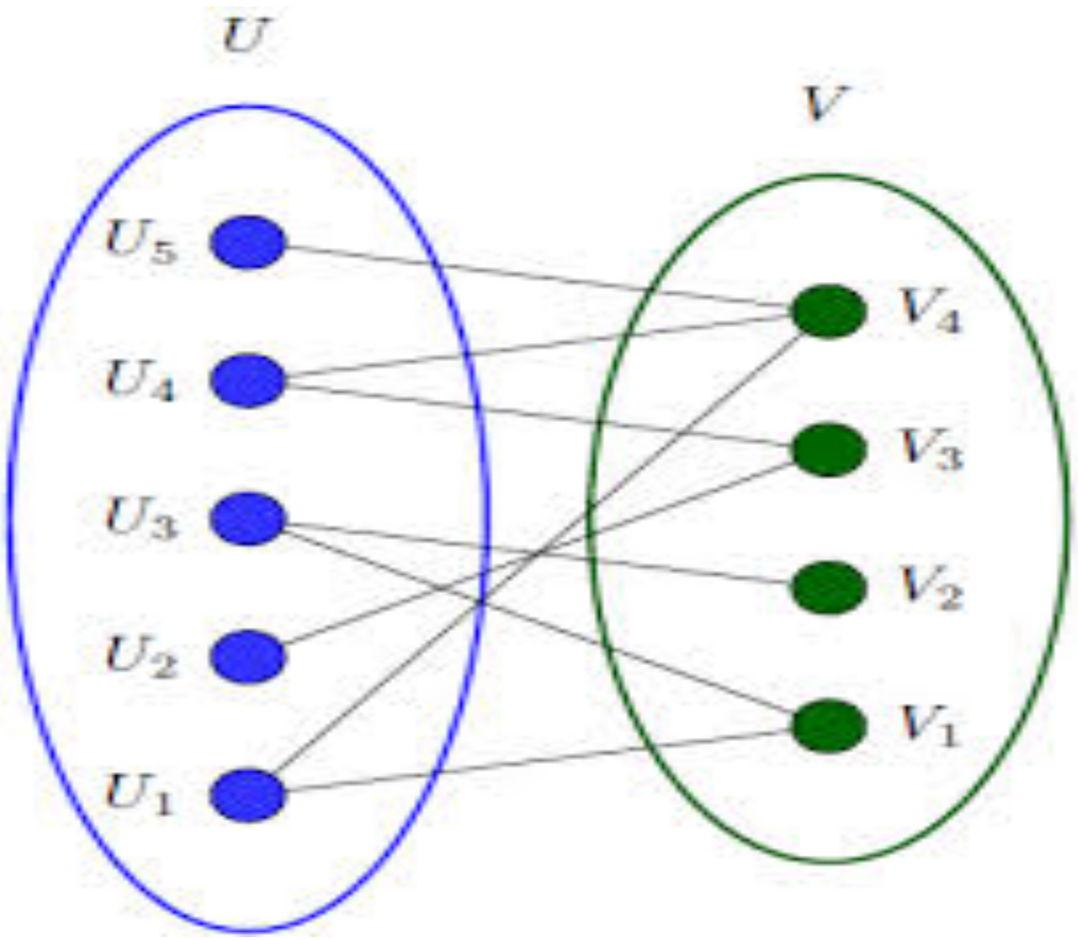
Business connections



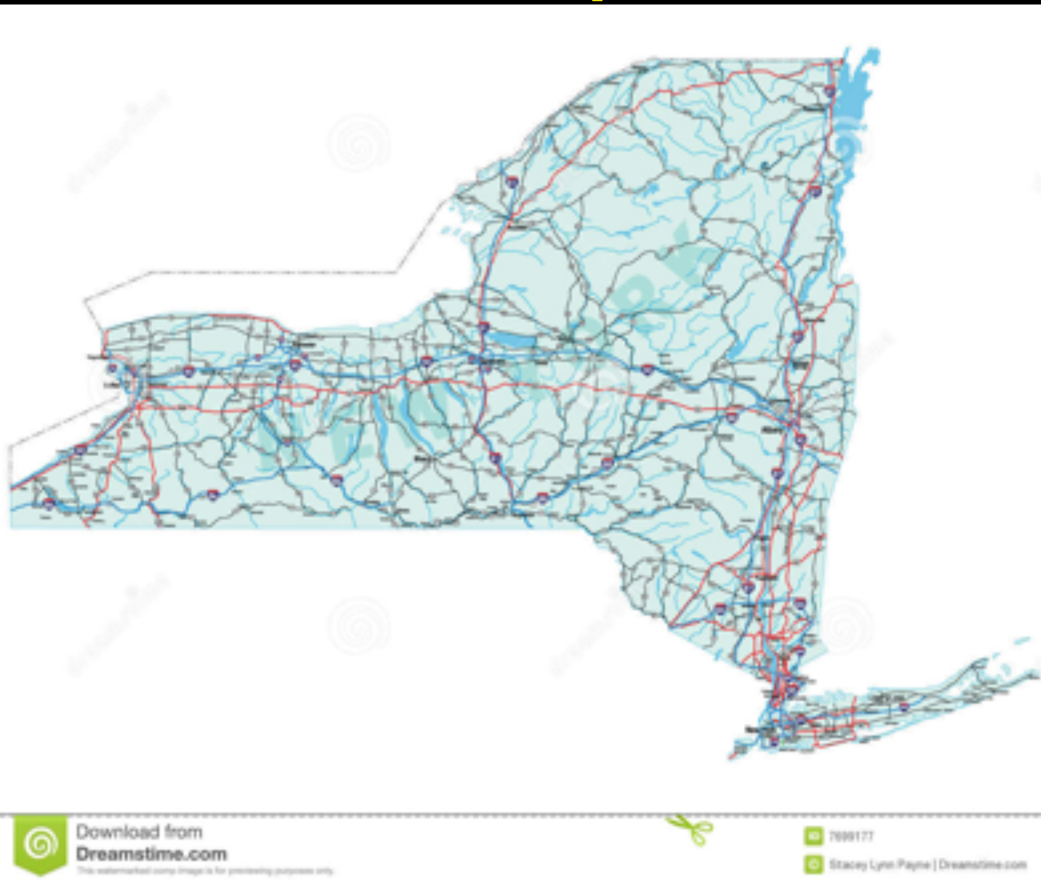


- Sharing customers
- Supplier relations

Bi-partite networks

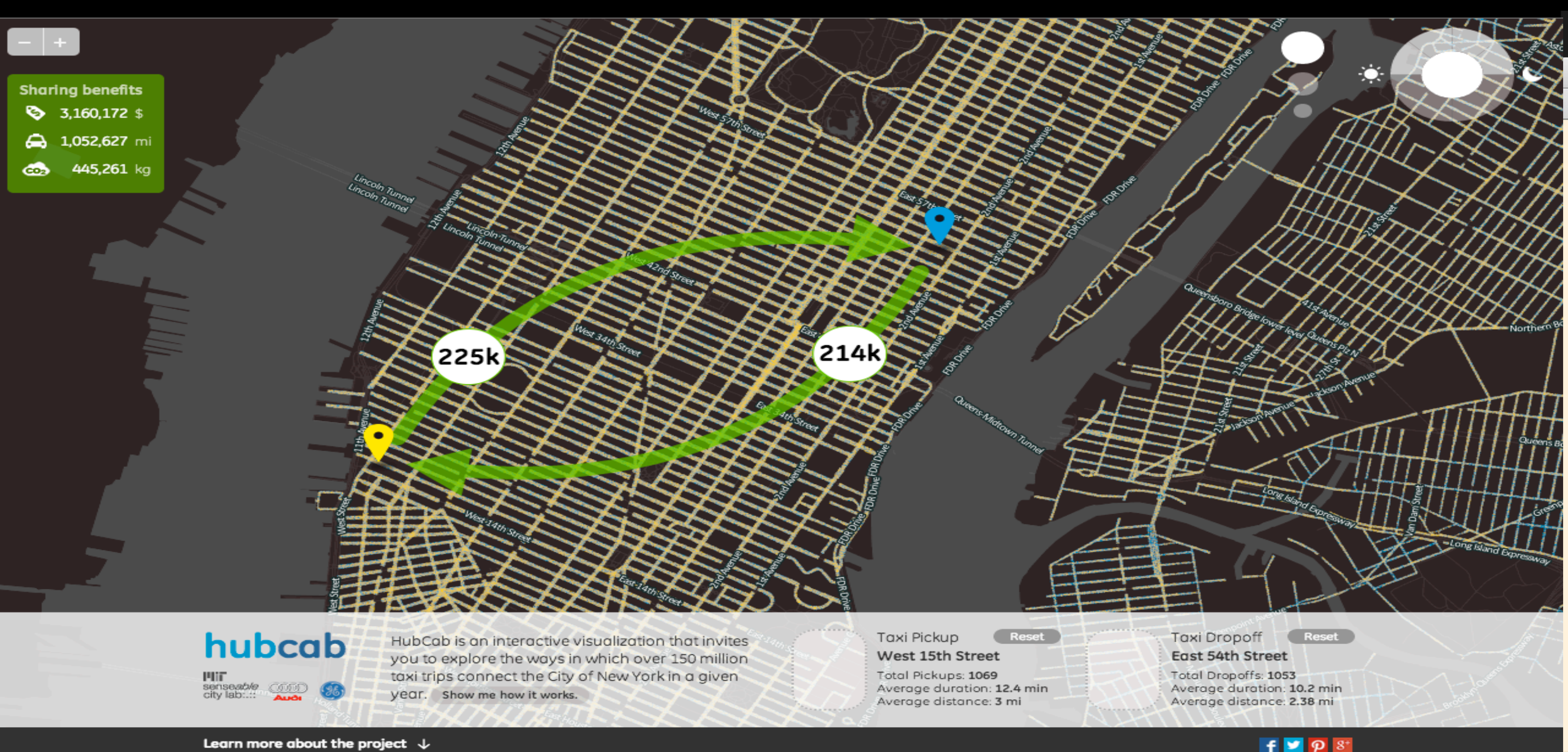


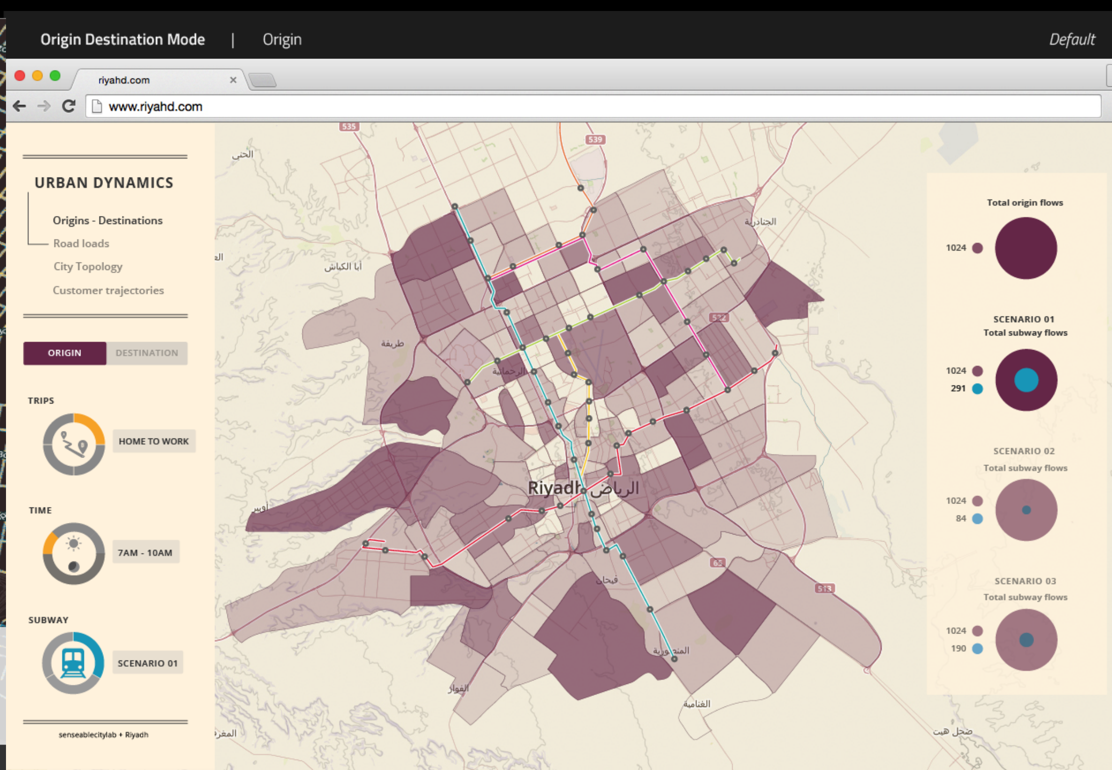
Urban transportation



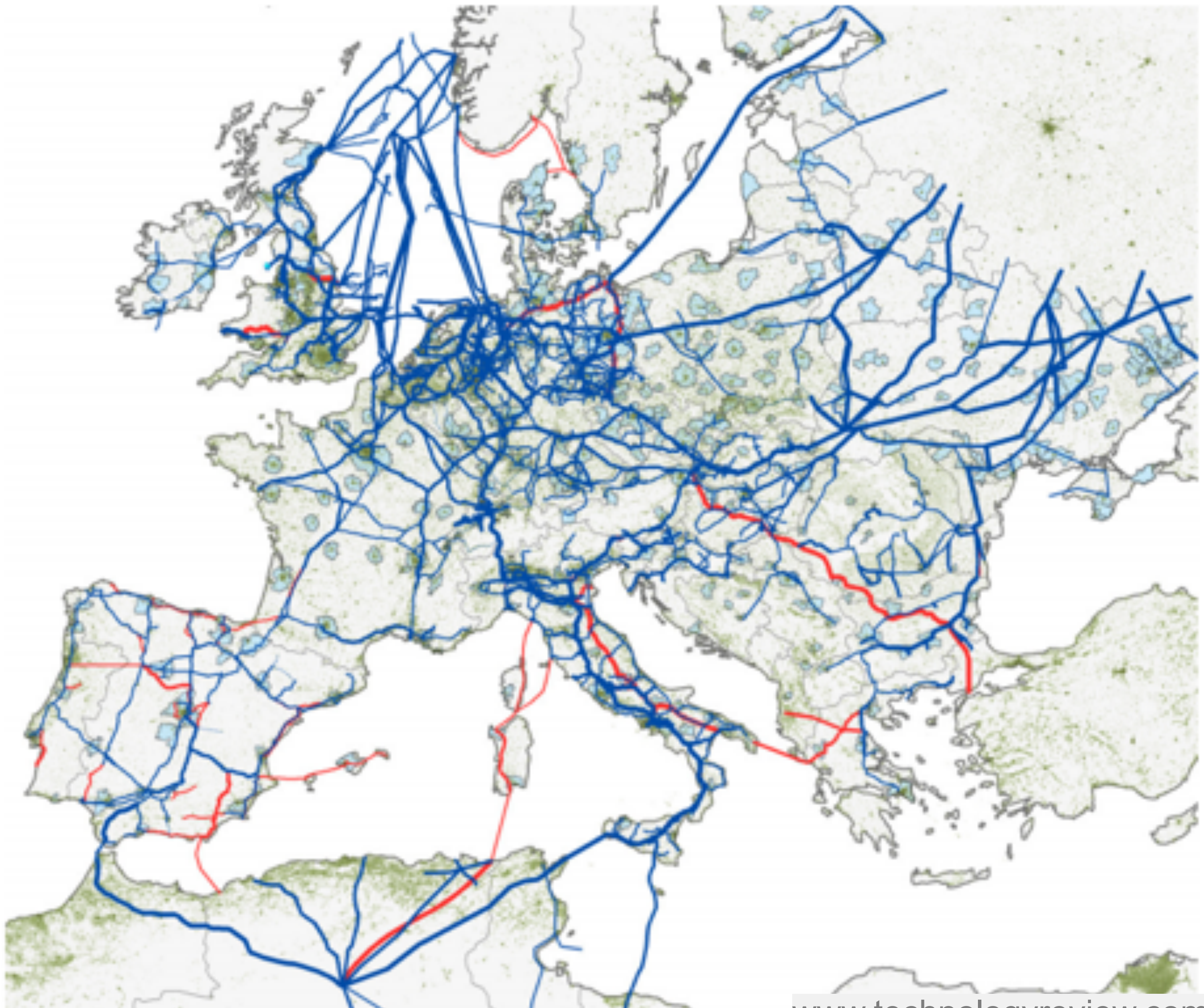
road network

Bus/subway/train routes

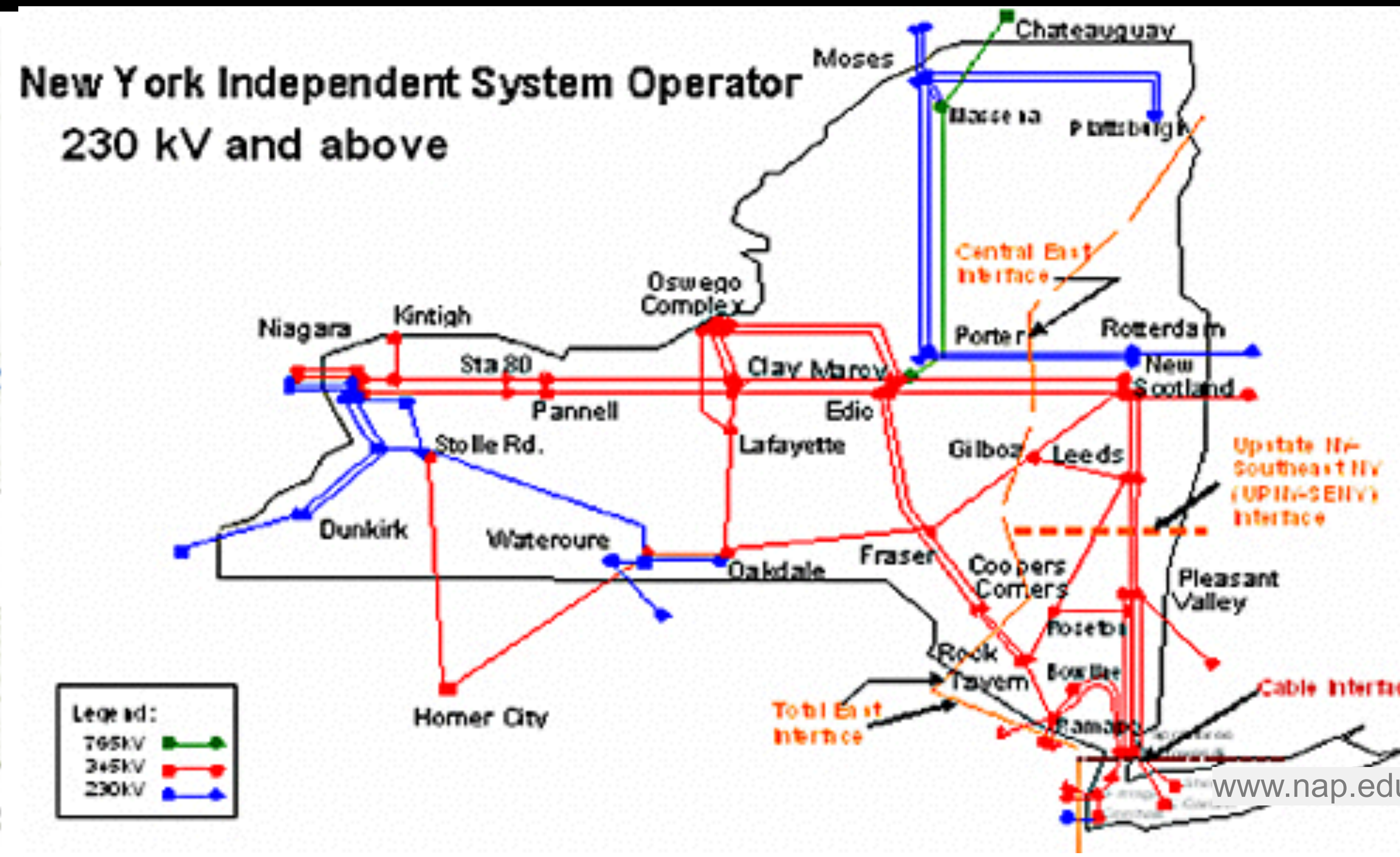




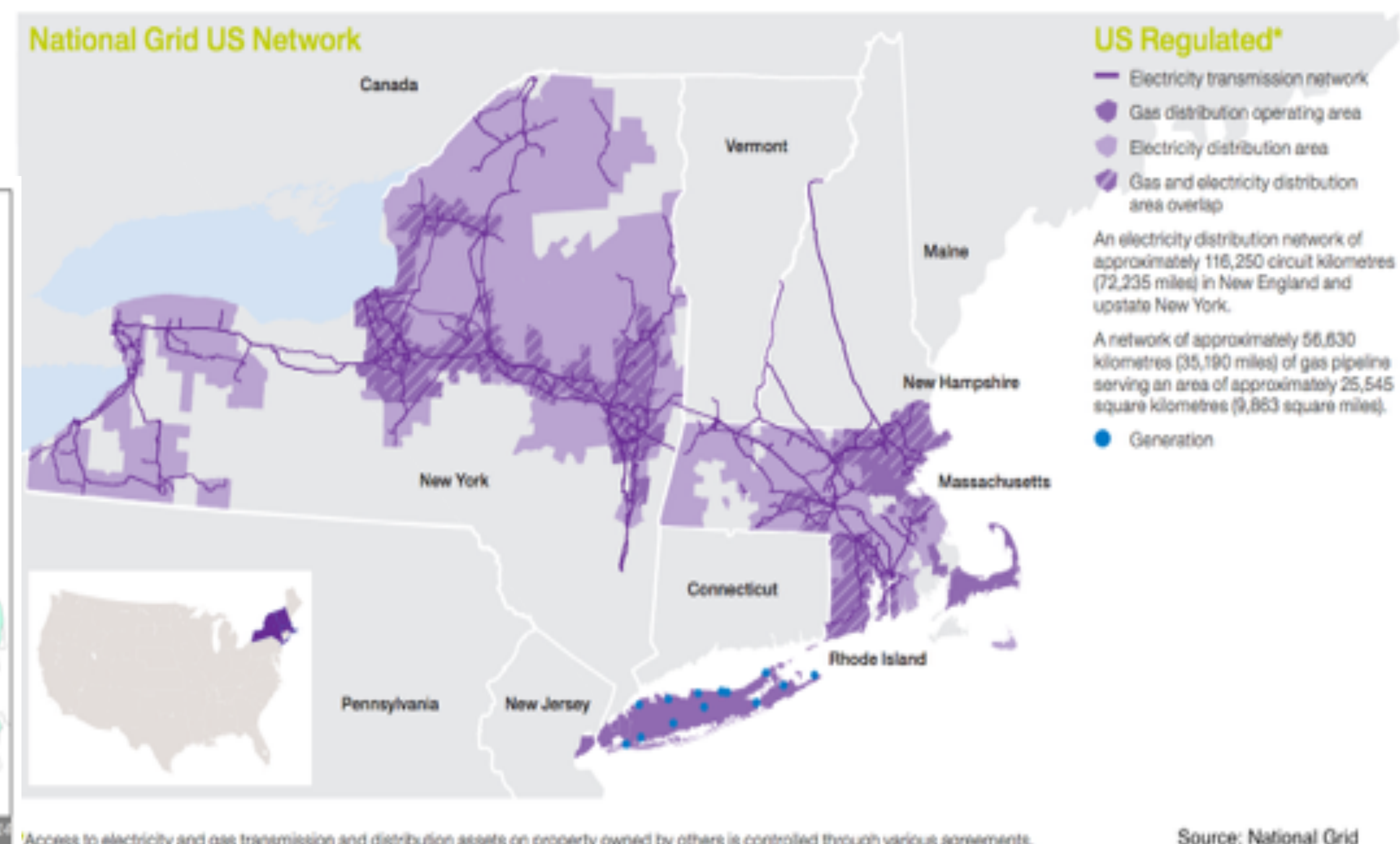
Other infrastructural networks



www.technologyreview.com

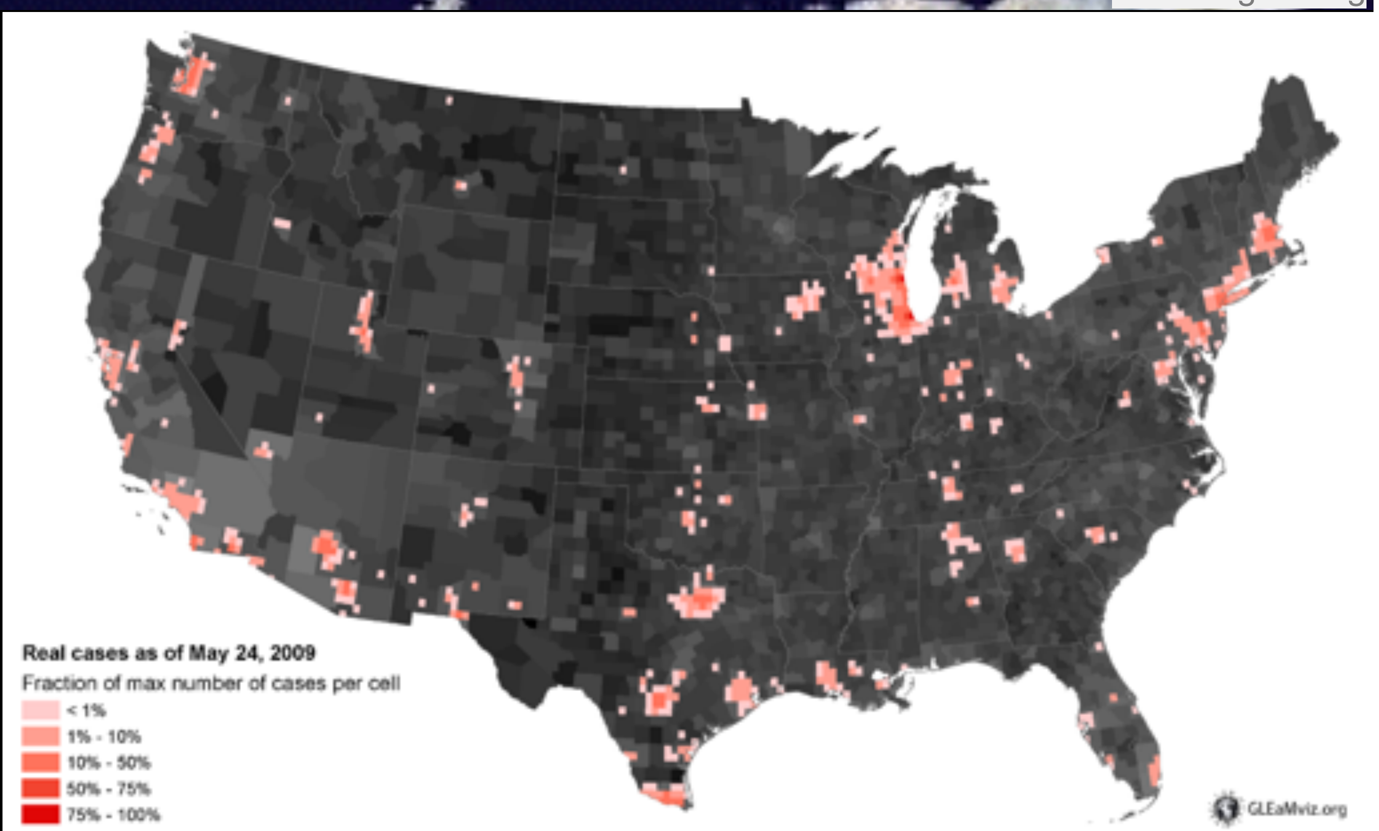


www.nap.edu

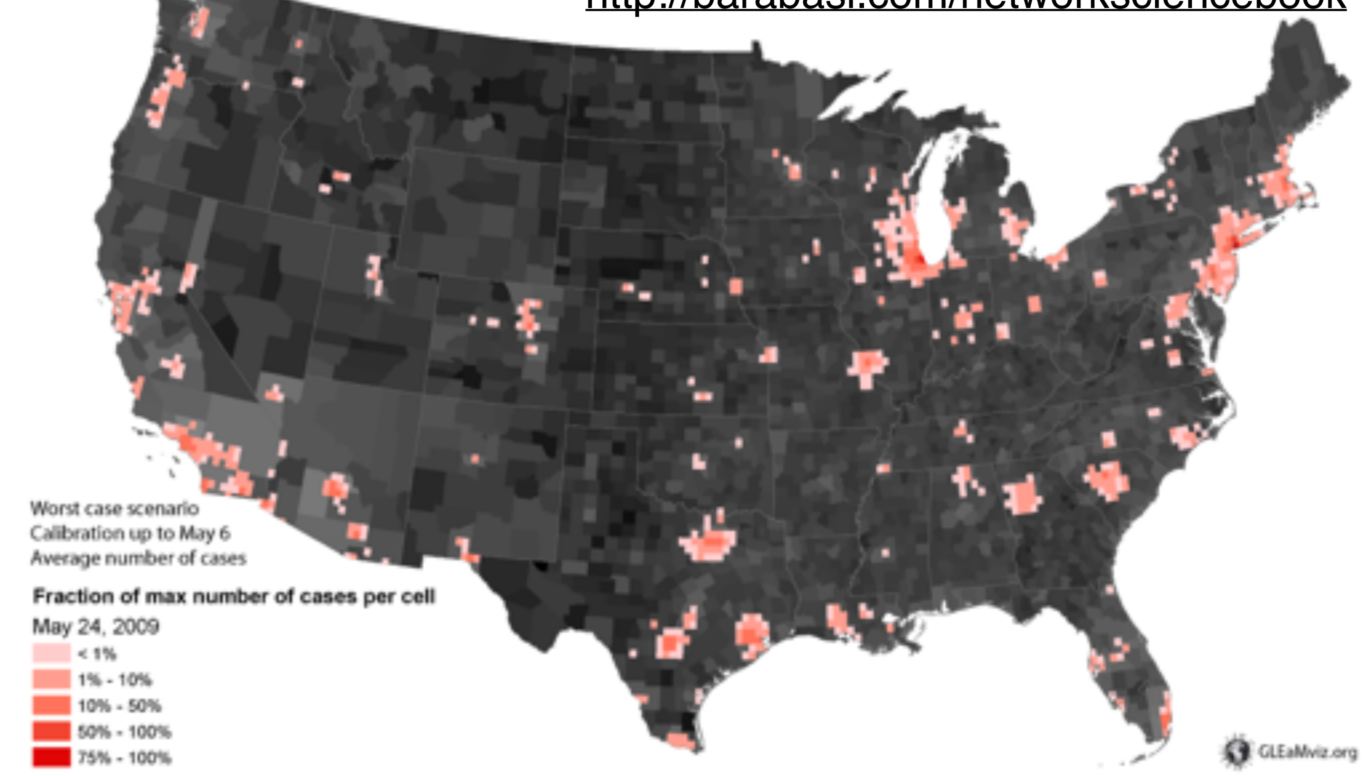


The Brooklyn Sewer System in 1893

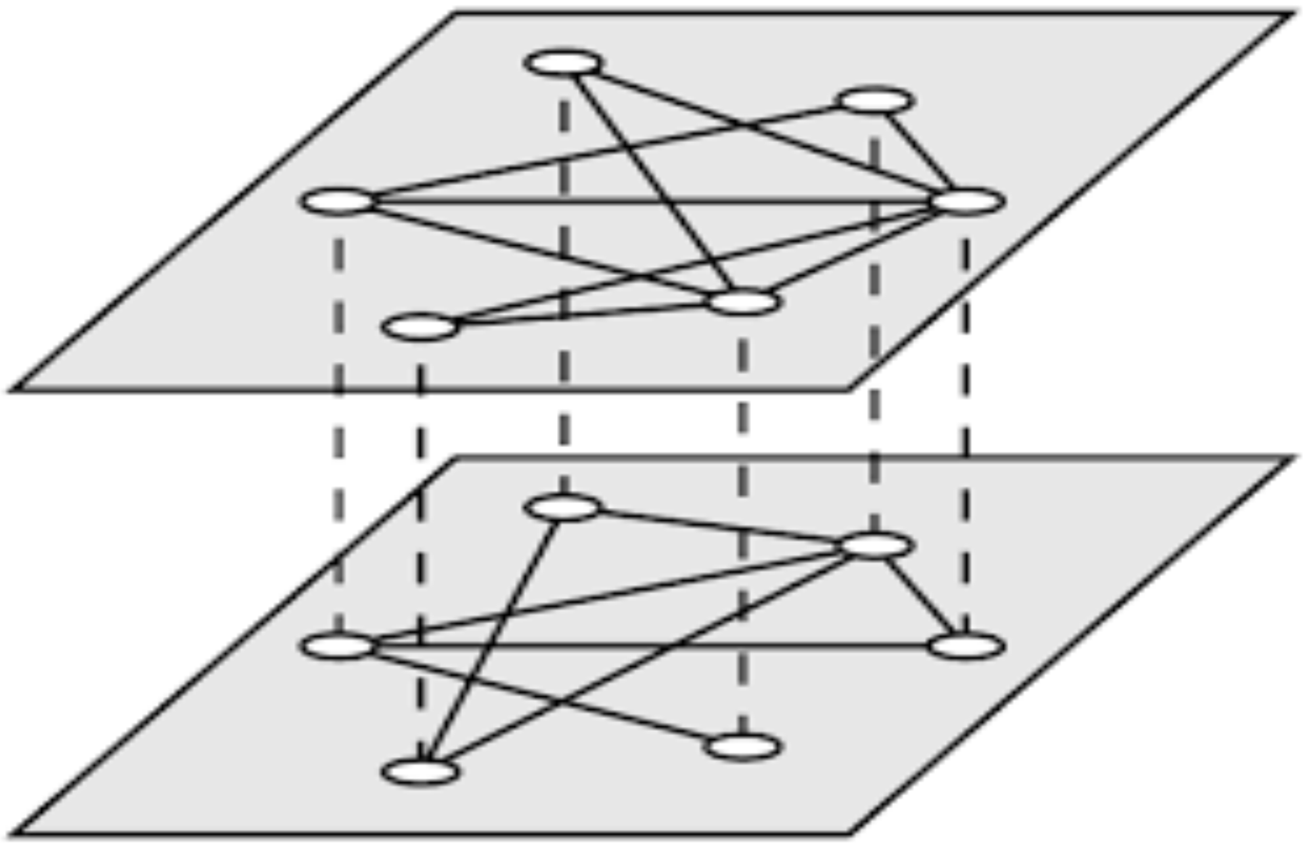
Epidemic networks

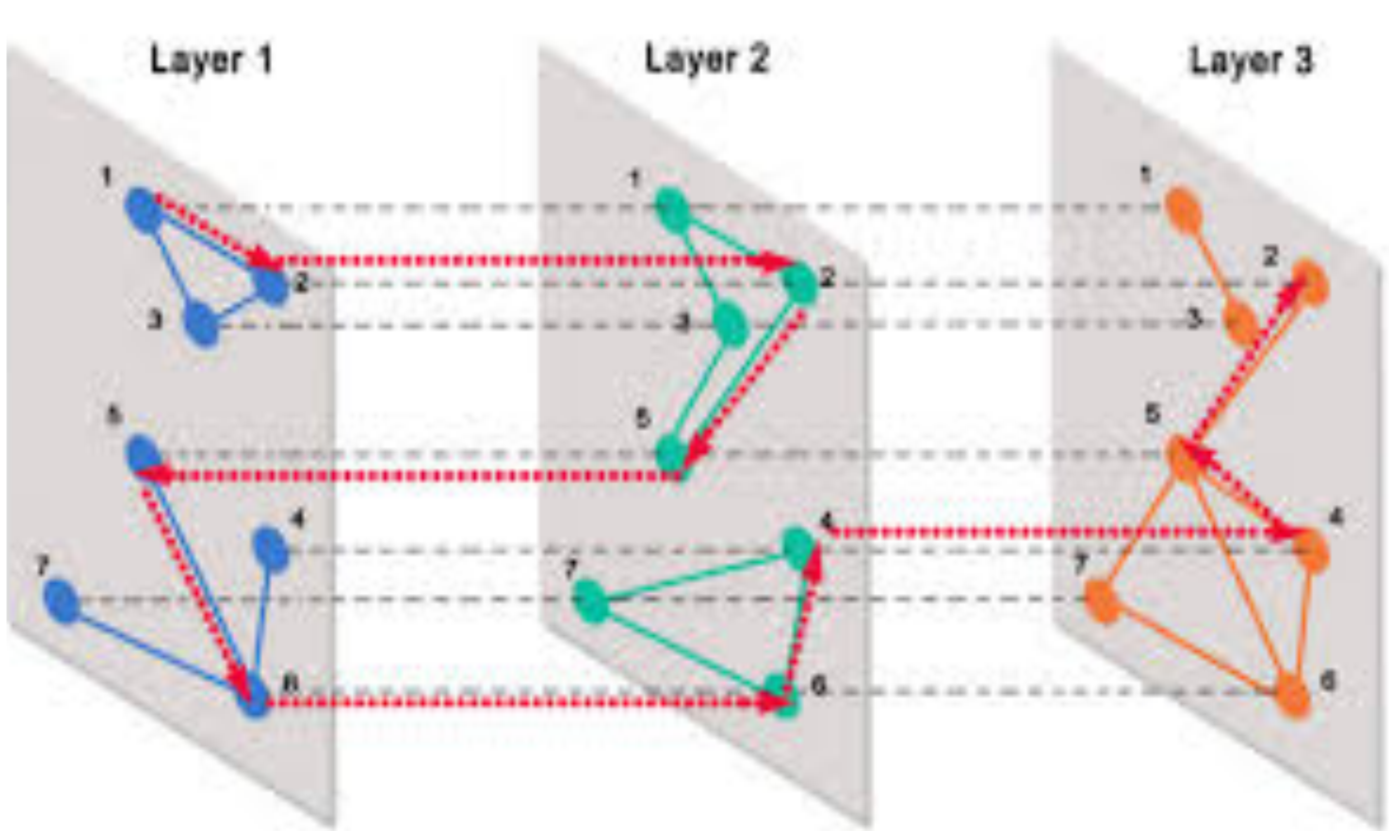


http://barabasi.com/networksciencebook



Multilayer/multiplex networks





LAYERS OF URBAN NETWORK

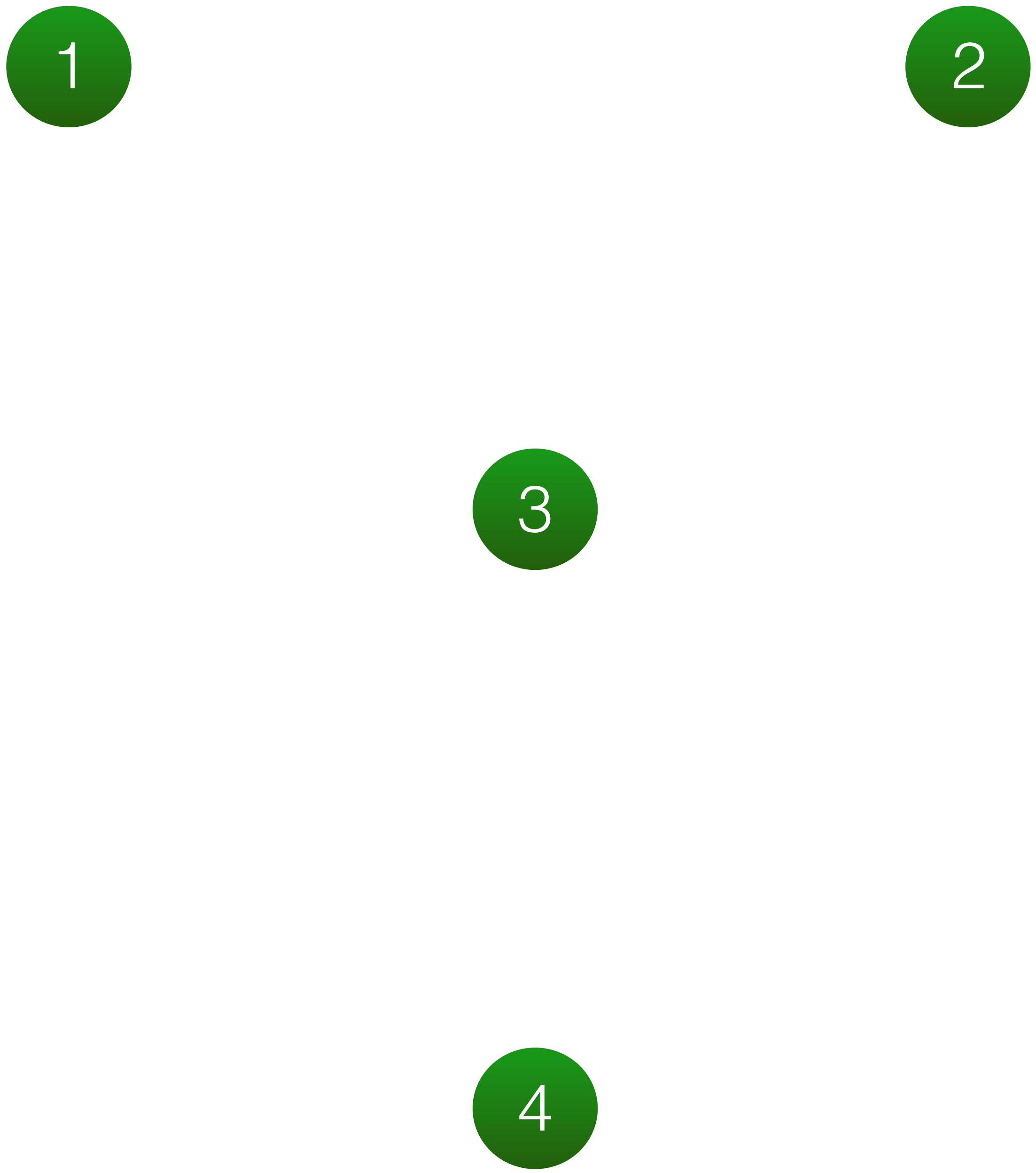
NETWORKS

- Human dynamics
- Traffic and road infrastructure
- Human interactions and social networks
- Professional interactions, collaboration
- Utility infrastructure: energy grid, water supply etc
- Waste removal
- Trade; goods supply and logistics
- Individual purchases
- Disease spread
- Knowledge propagation and innovation

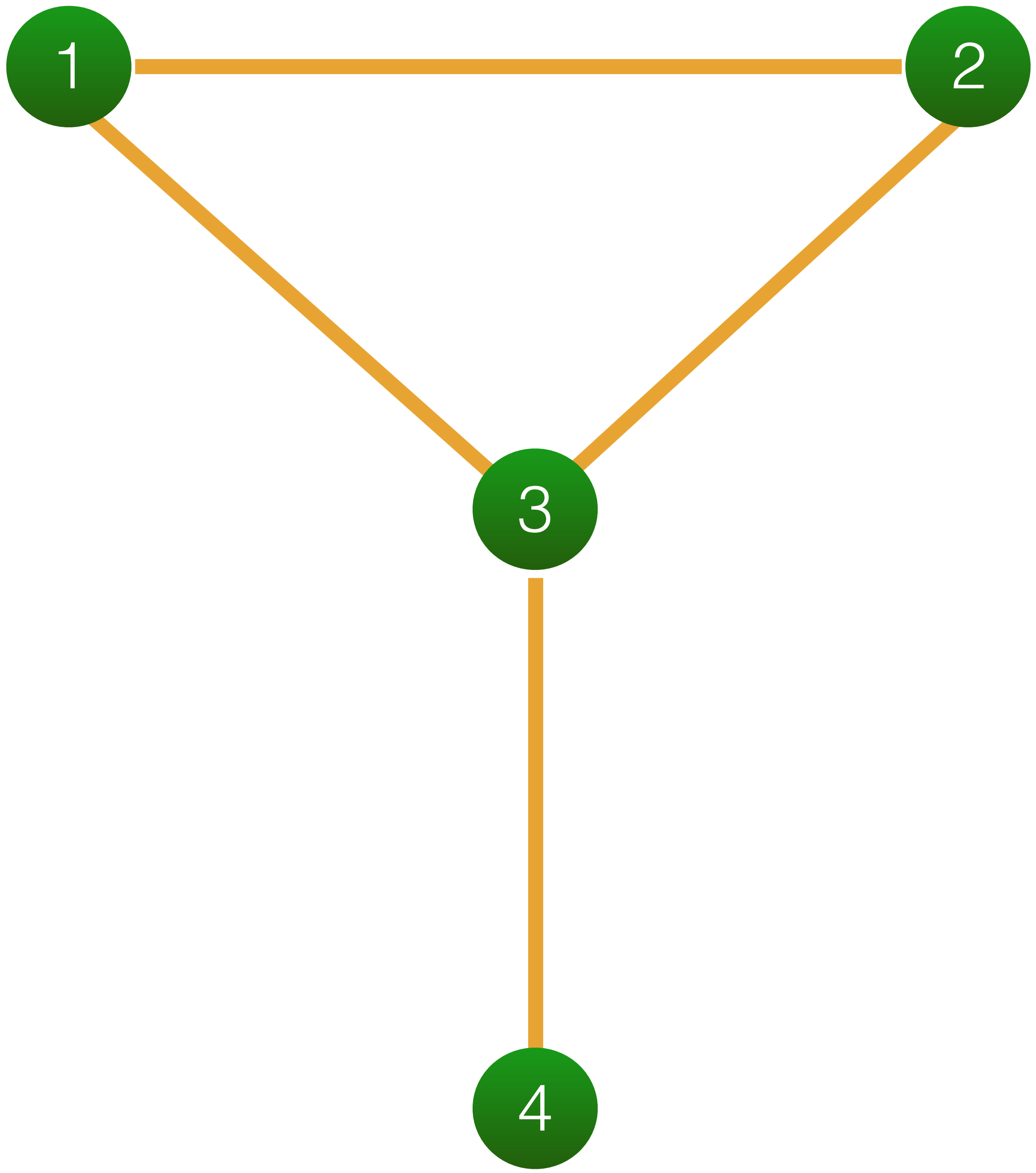
Applications of urban network analysis

- Urban dynamics: estimating actual current local population and its structure;
- Urban epidemiology;
- Urban transportation:
  - Routing,
  - planning and optimizing public transportation, taxies
  - traffic control
  - planning parking spaces
  - car sharing
  - predicting load on the urban transportation system, reaction to shocks and system resilience
- Regional delineation;
- Delineating city boundaries;
- Land-use classification;
- Planning and estimating resilience of urban infrastructure;
- Revealing social interactions in local communities
- Estimating economic health and potential of local neighborhoods
- Revealing important event and happenings though urban network disruptions, and predicting potential impact

Concept of a network

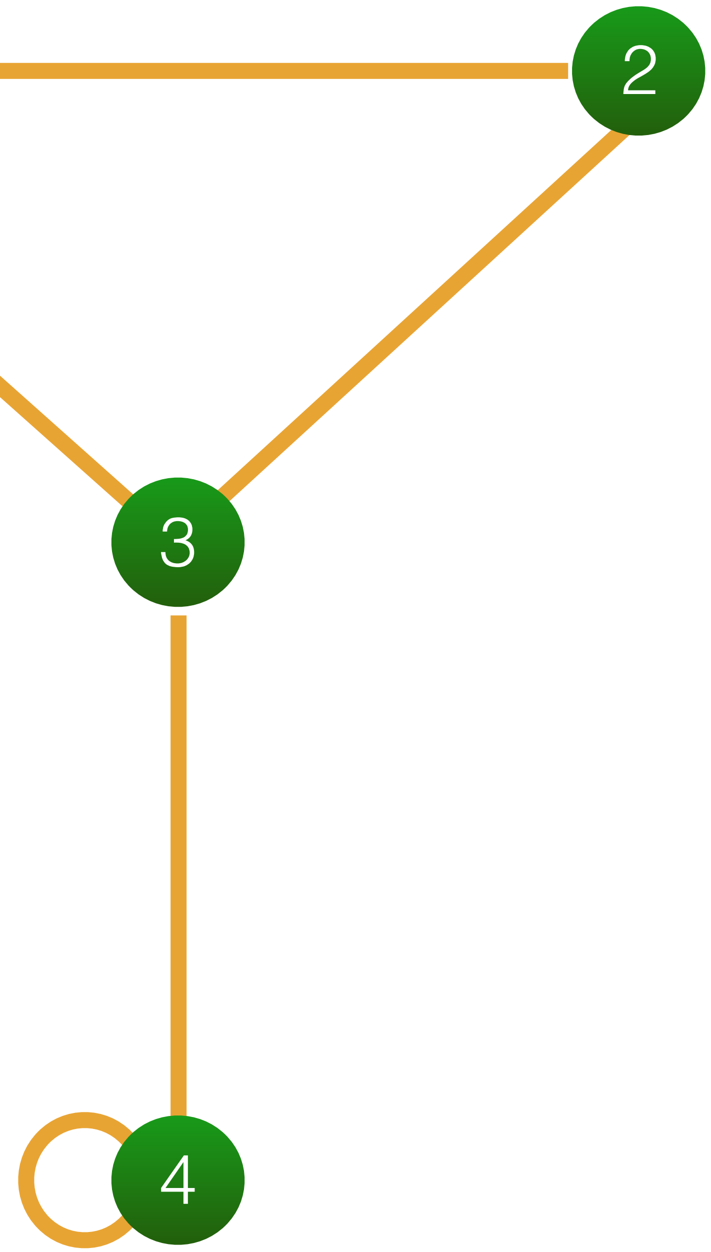


Concept of a network

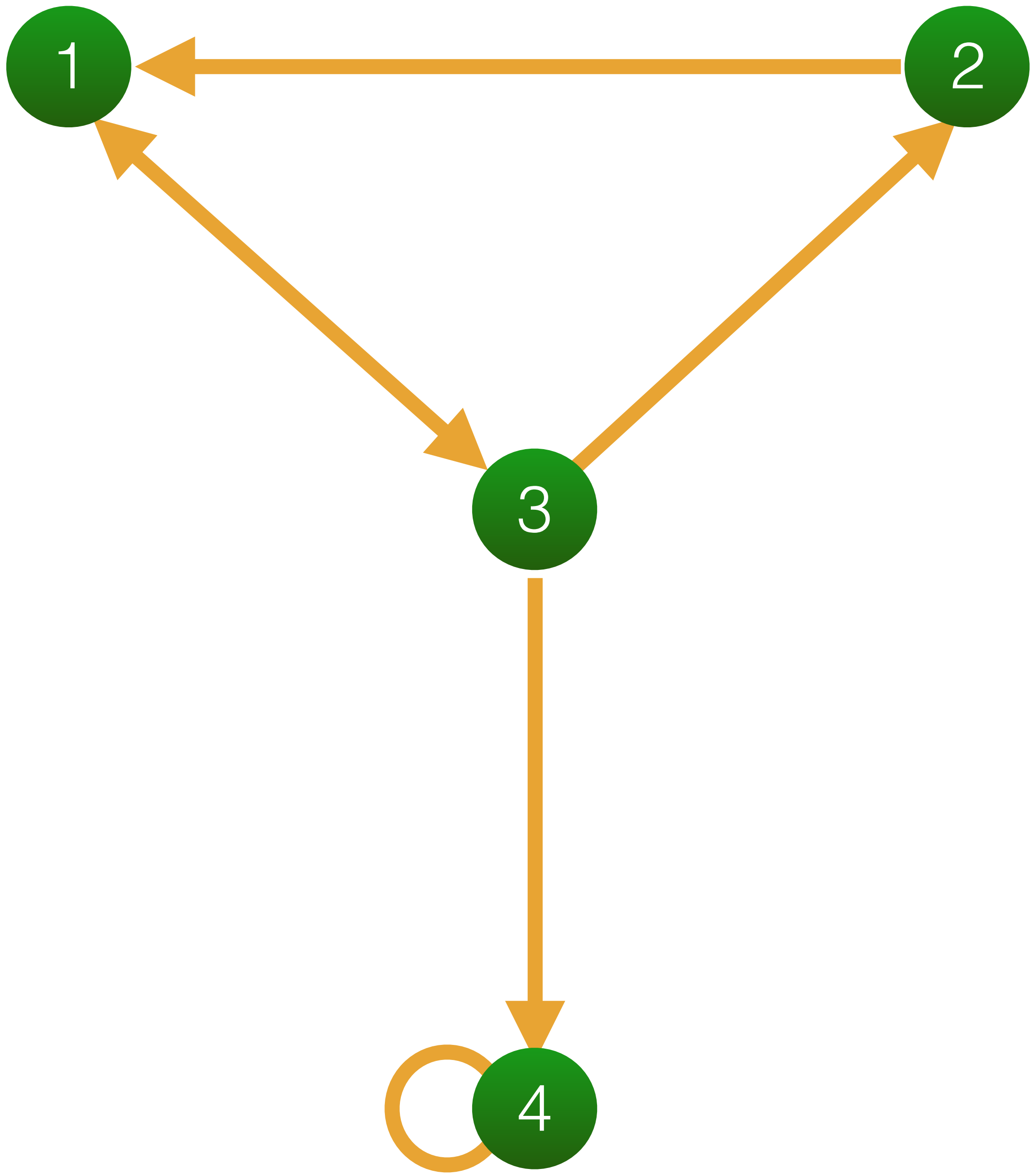


$$G = (N, E)$$

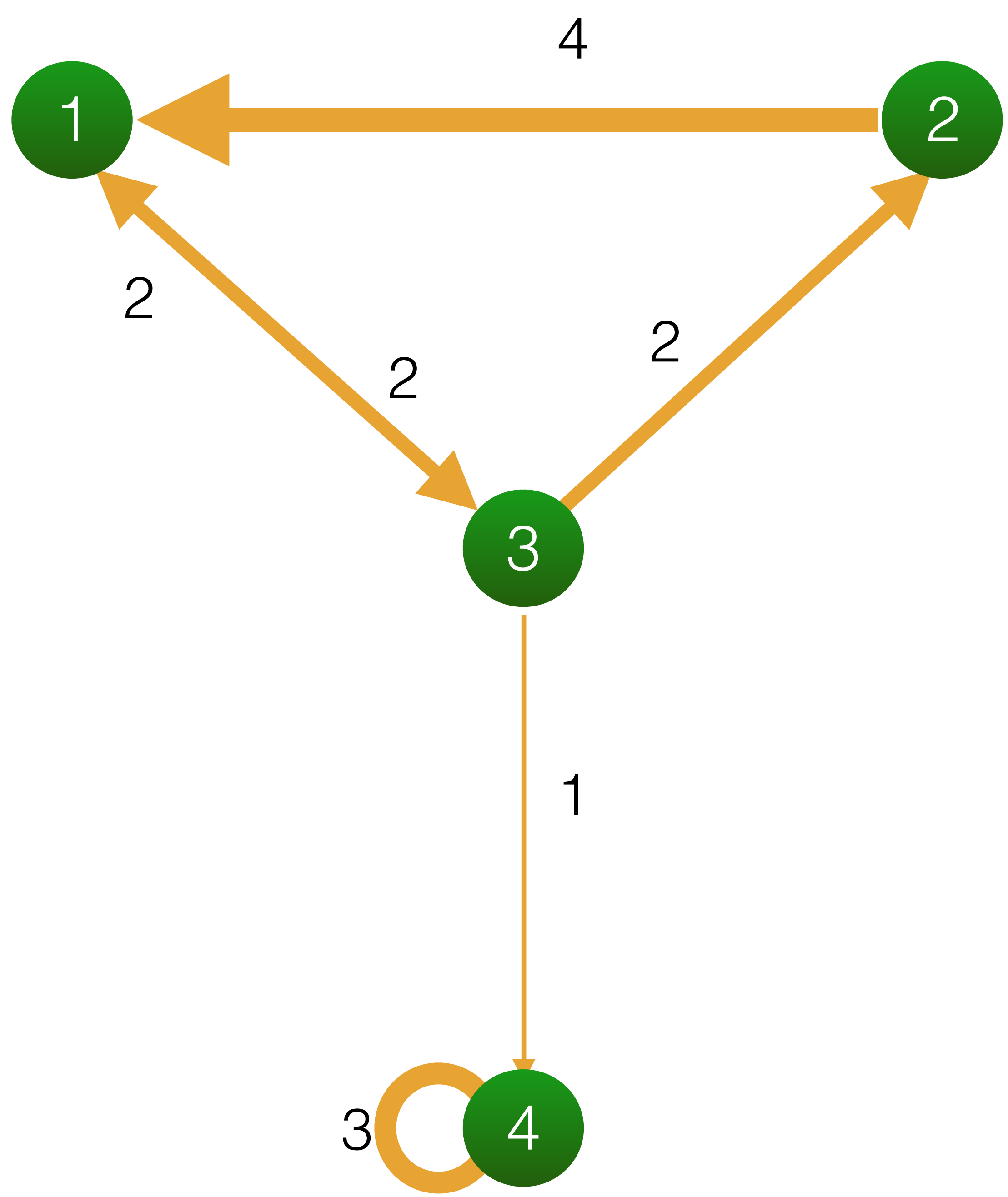
Concept of a network



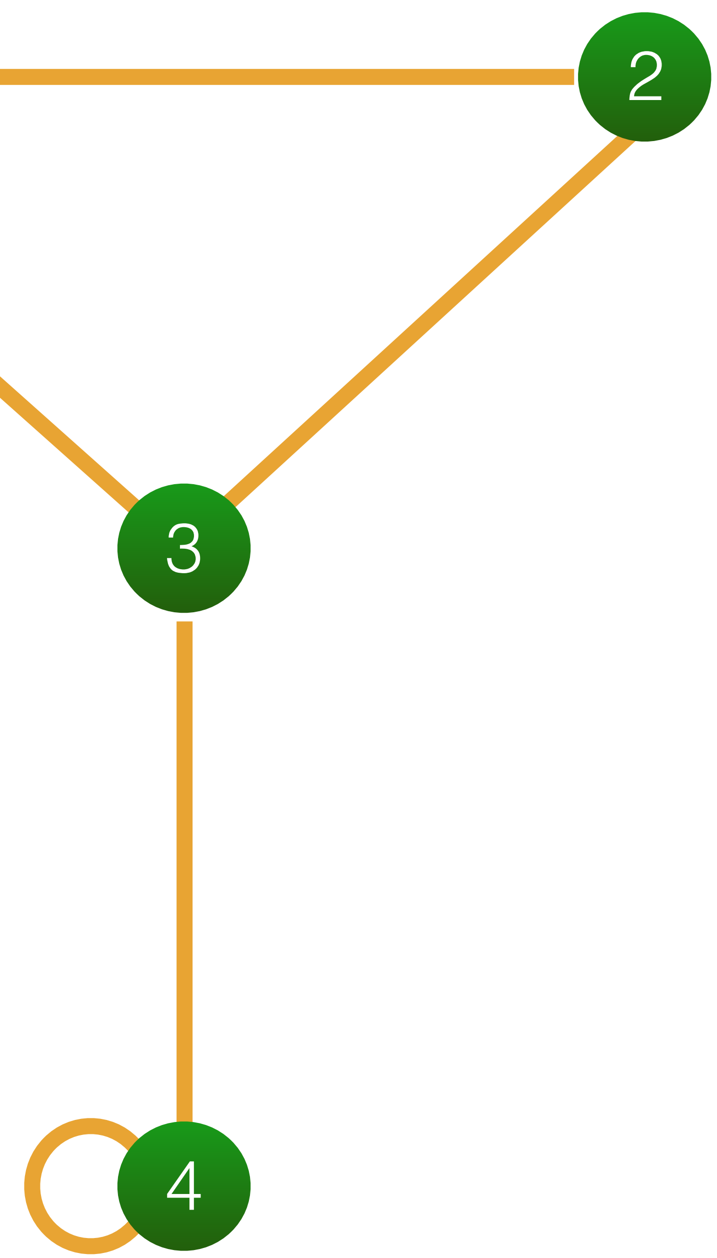
Concept of a network



Concept of a network



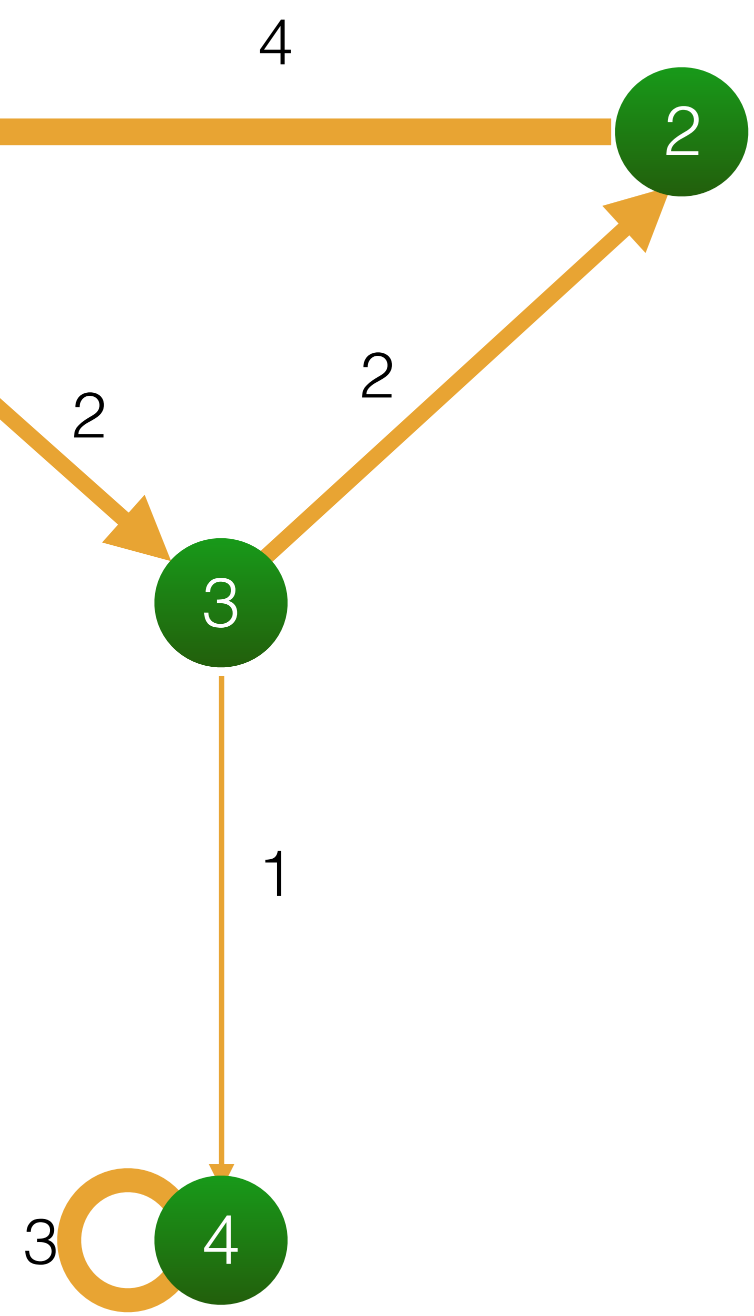
Concept of a network



O	D
1	2
1	3
2	2
3	4
4	4

	1	2	3	4
1	0	1	1	0
2	1	0	1	0
3	1	1	0	1
4	0	0	1	1

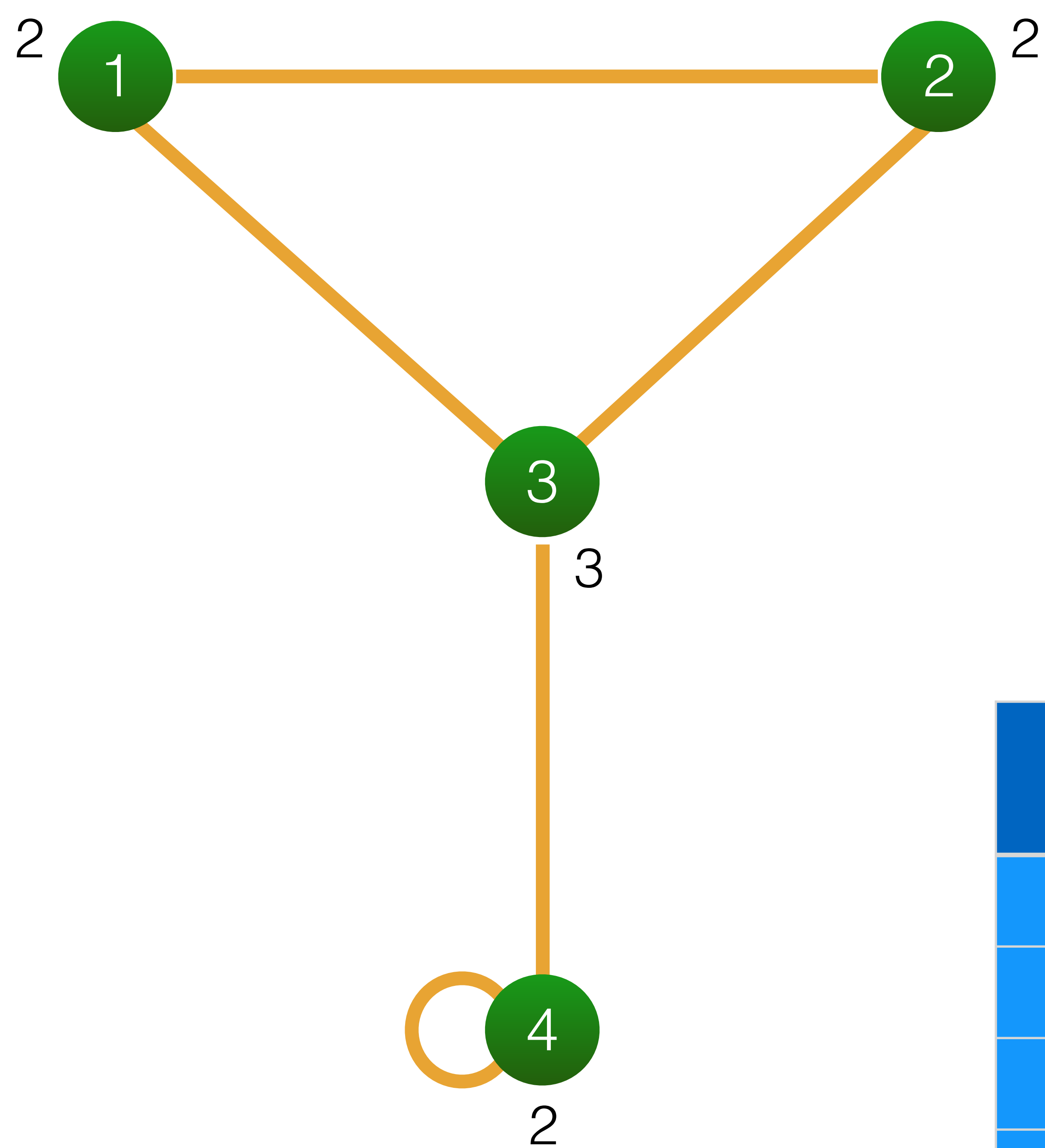
Concept of a network



O	D	W
2	1	4
1	3	2
3	1	2
3	2	2
3	4	1
4	4	3

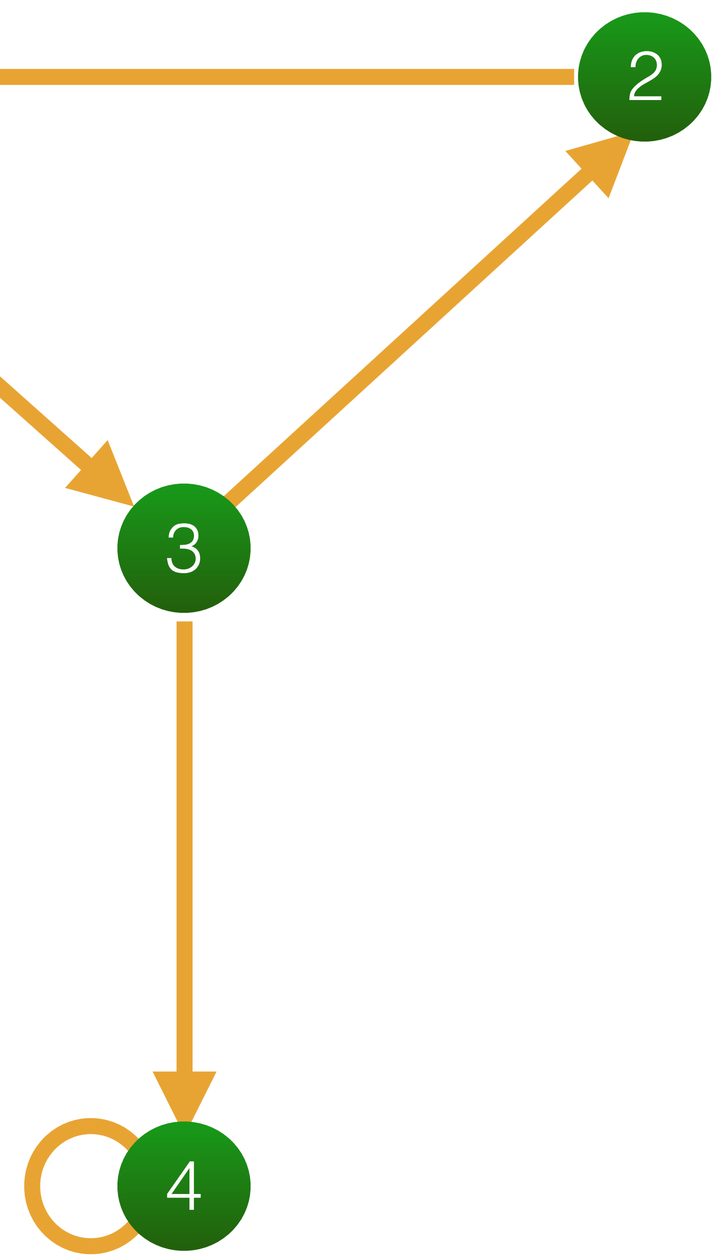
	1	2	3	4
1	0	0	2	0
2	4	0	0	0
3	2	2	0	1
4	0	0	0	3

Node degree



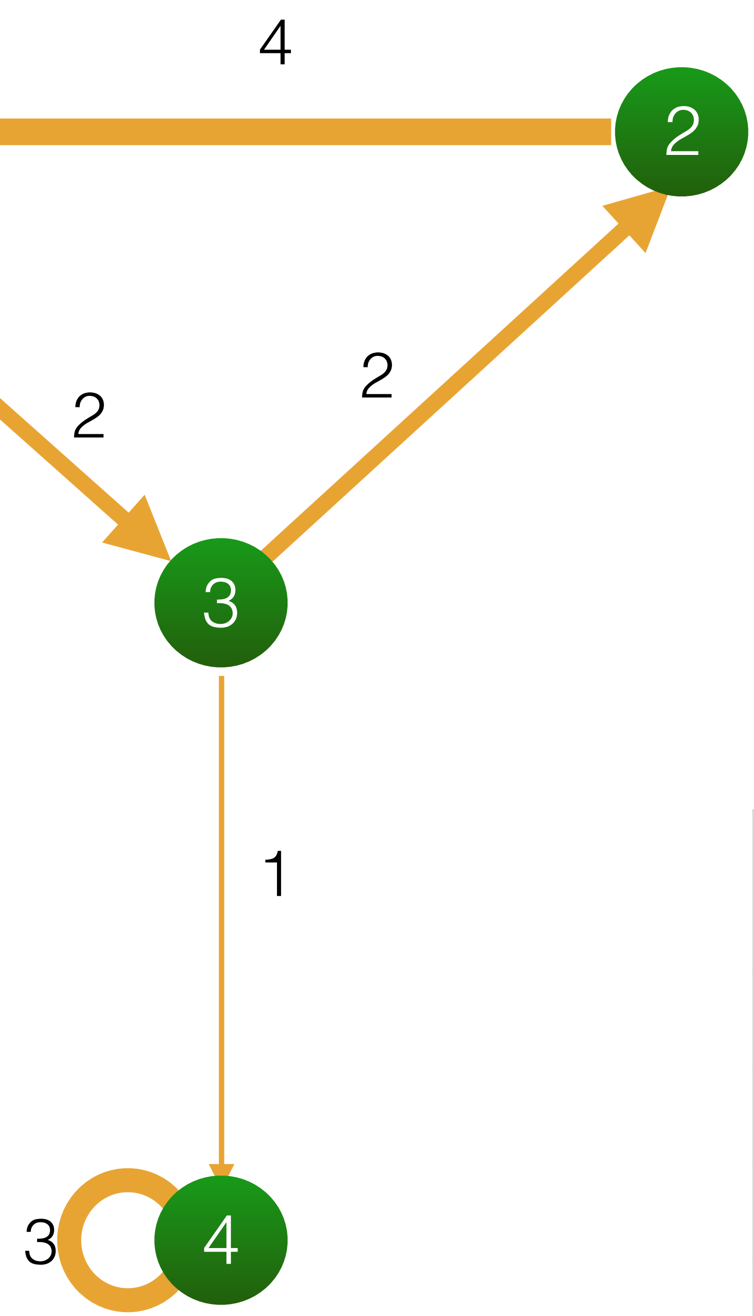
	1	2	3	4	D
1	0	1	1	0	2
2	1	0	1	0	2
3	1	1	0	1	3
4	0	0	1	1	2

Directed network: node in/out degree



	1	2	3	4	Out
1	0	0	1	0	1
2	1	0	0	0	1
3	1	1	0	1	3
4	0	0	0	1	1
In	2	1	1	2	6

Weighted network: node strength



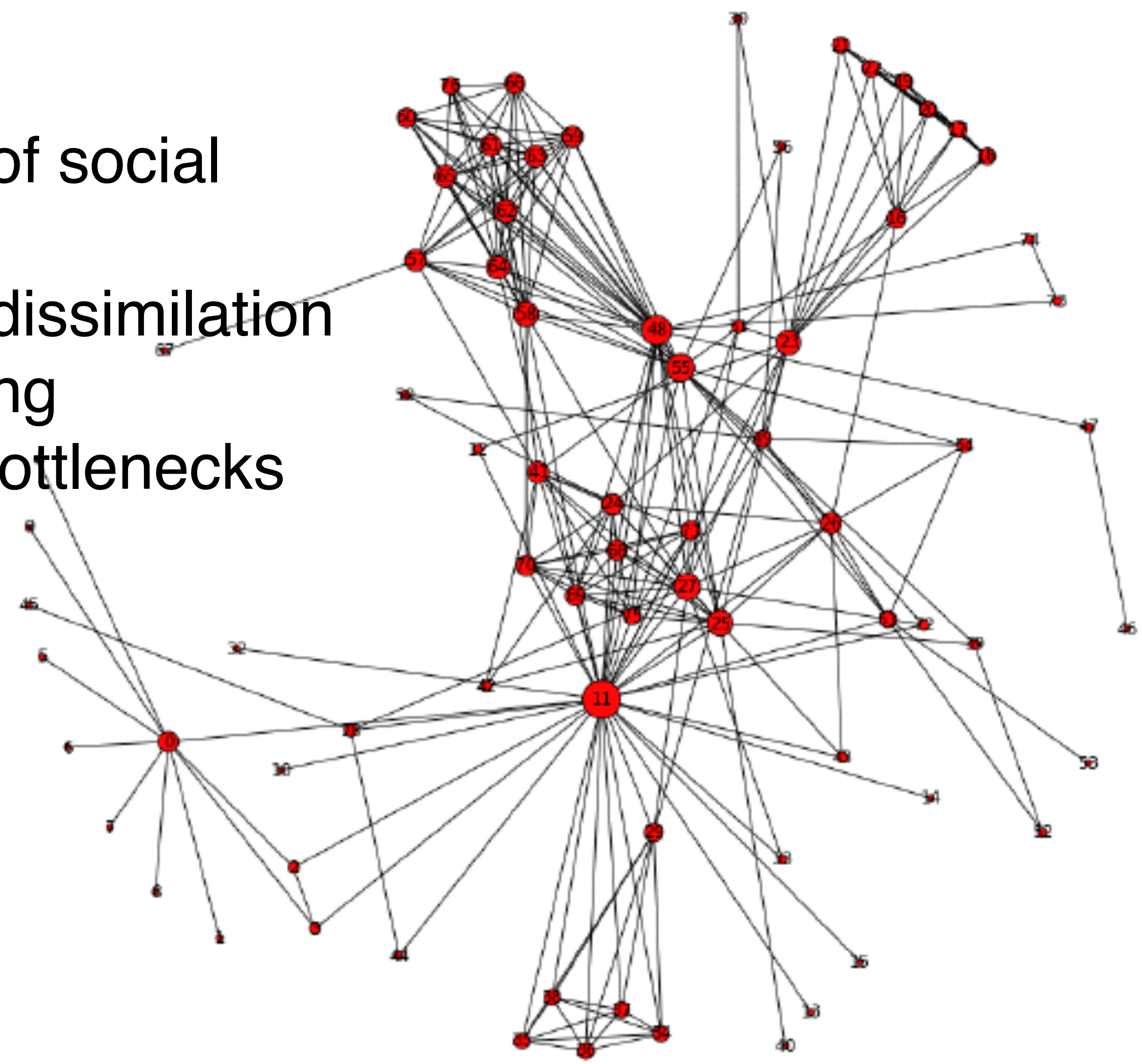
	1	2	3	4	Out
1	0	0	2	0	2
2	4	0	0	0	4
3	2	2	0	1	5
4	0	0	0	3	3
In	6	2	2	4	14

Node centrality

How important the node is for the network?

Why do we care?

- detecting influential members of social networks
- strategy for faster information dissimilation
- epidemic prevention and fighting
- network resilience, detecting bottlenecks
- web-search - important pages

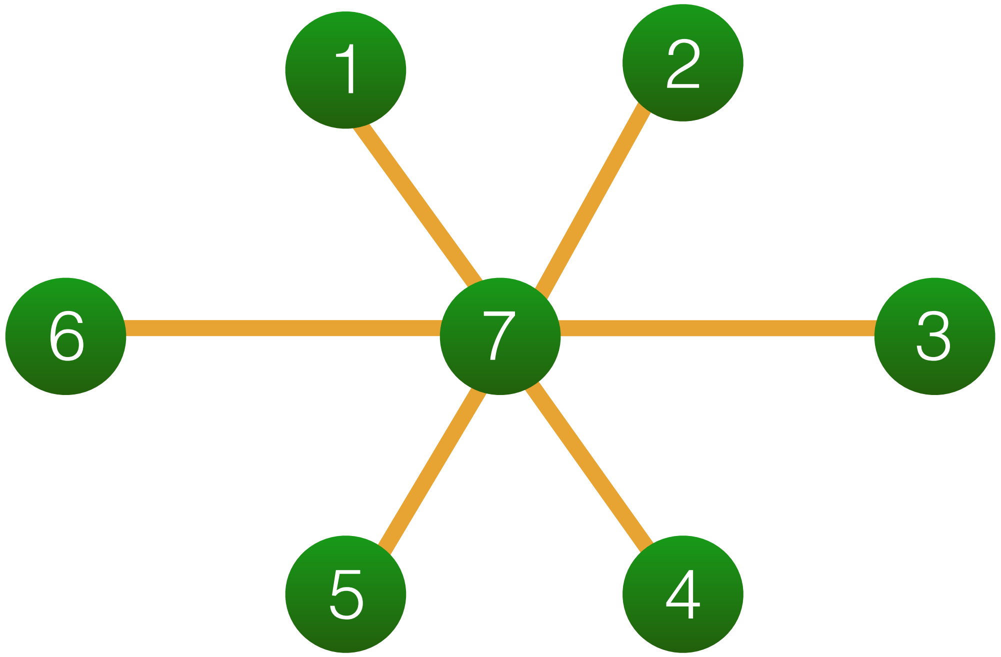


Node centrality

Many answers to this question, depending on what do we mean by importance

Sometimes it’s obvious - degree

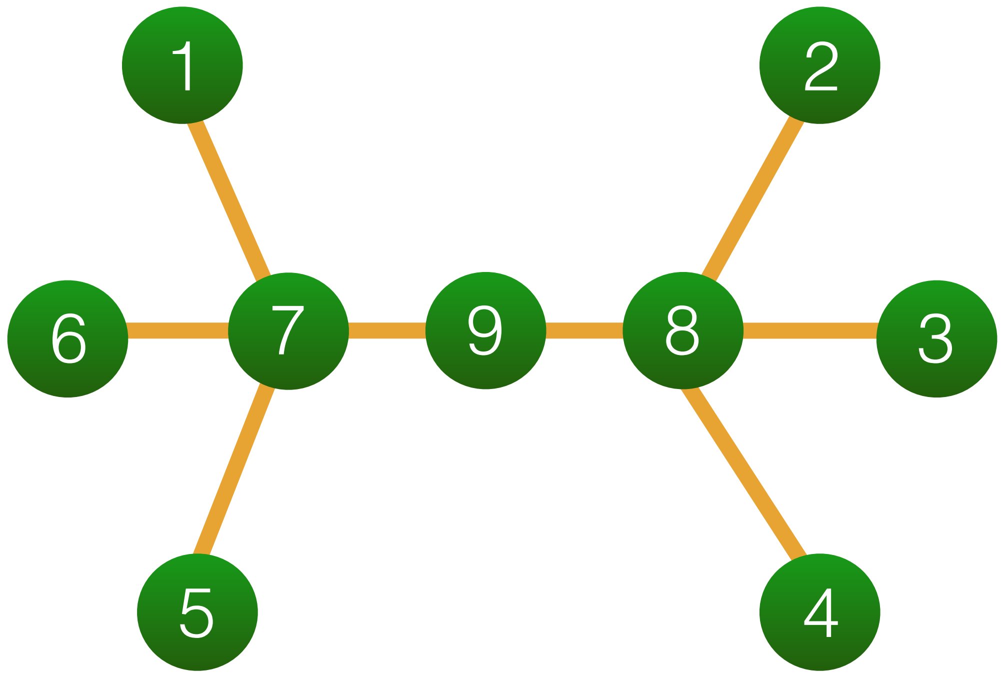
Sometimes degree is insufficient



Scott Adams: the power a person holds in the organization is inversely proportional to the number of keys on his keyring. A janitor has keys to every office, and no power. The CEO does not need a key: people always open the door for him.

Node centrality

Sometimes degree is insufficient:



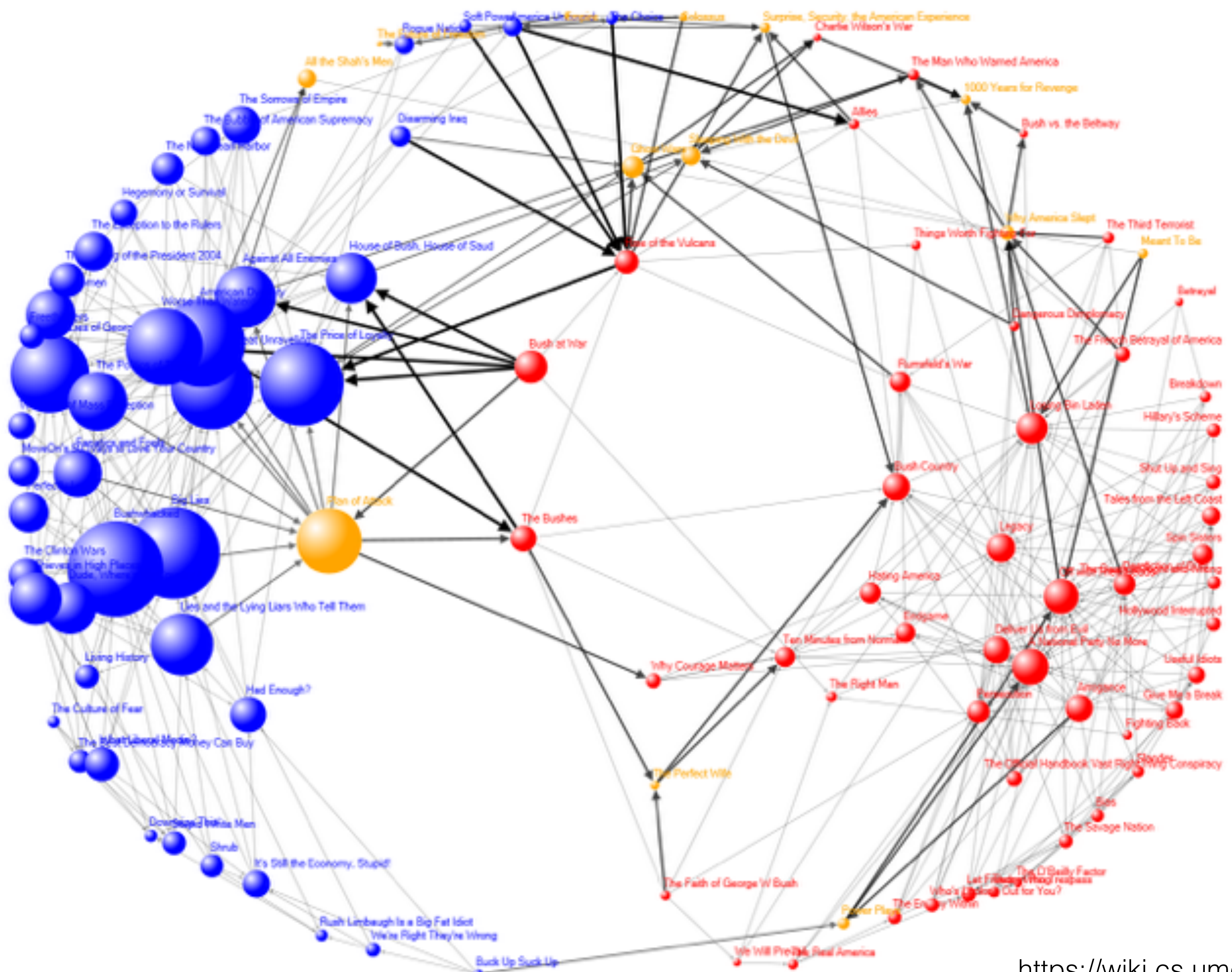
Bridges, brokerage between key parts

Betweenness centrality

Closeness centrality

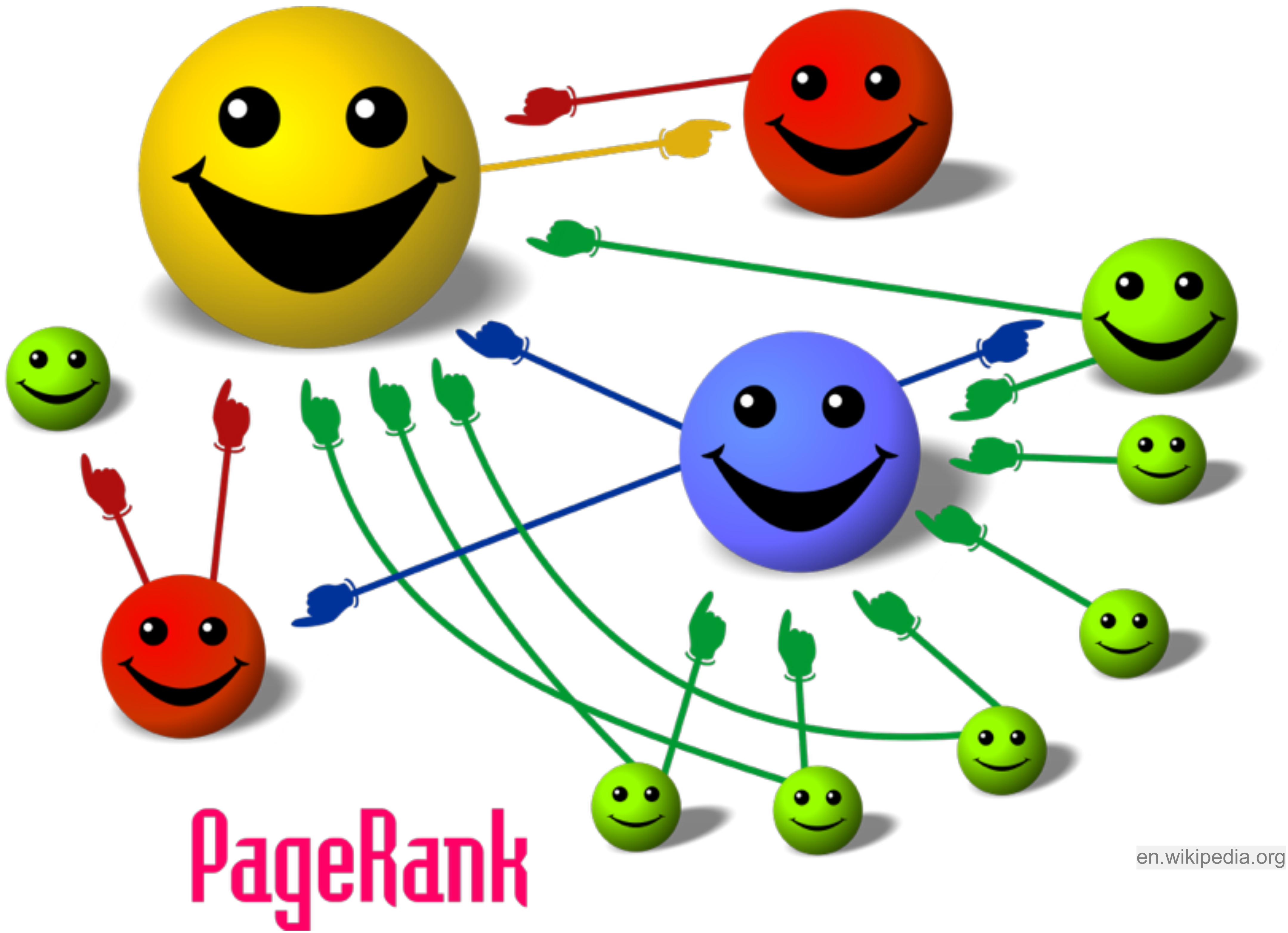
Node centrality

Eigenvector centrality family:



https://wiki.cs.umd.edu

Node centrality



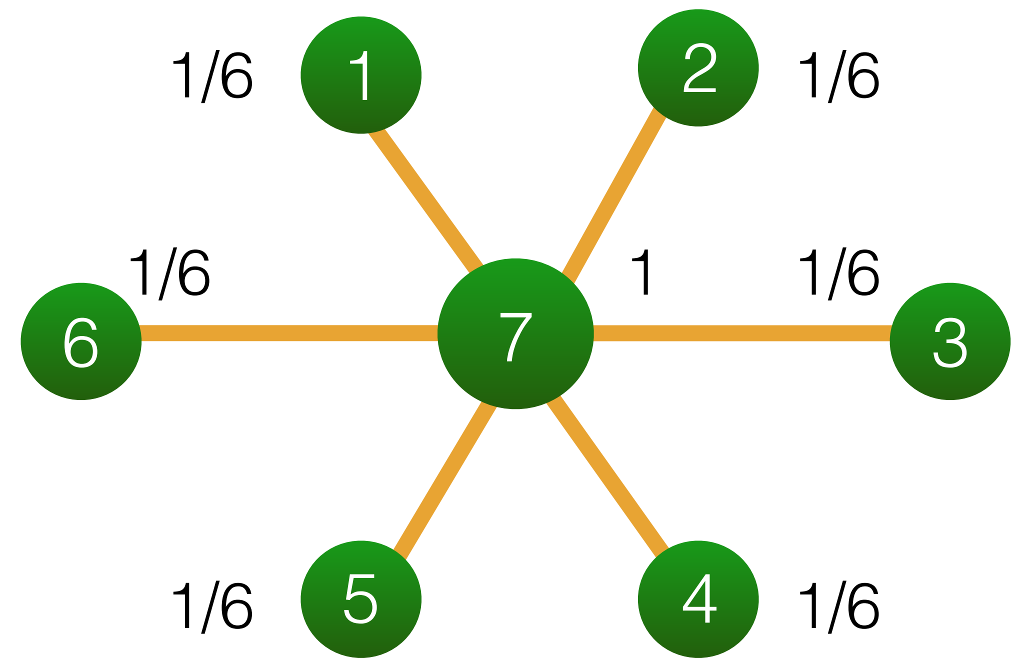
Degree centrality

$$N = \{a_i, i = 1..n\},\ E = \{(a_j, b_j)\}$$

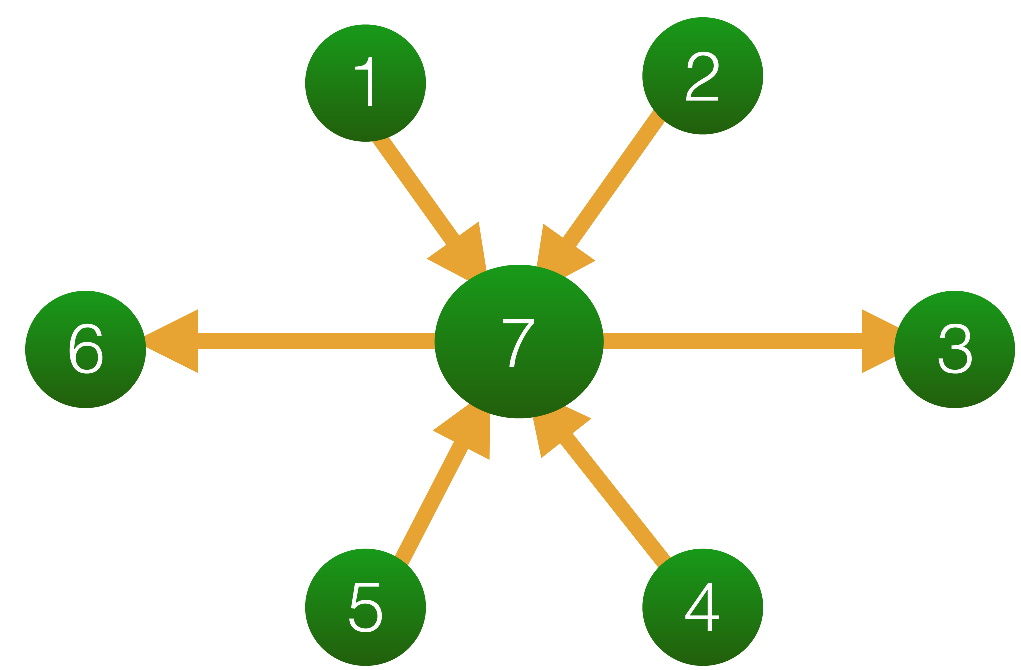
$$\mathrm{degree\_centrality}(a) = |\{(a,b) \in E\}|$$

$$global\ normalization : c(a) = \frac{degree\_centrality(a)}{\sum_b degree\_centrality(b)}$$

$$local\ normalization : c(a) = \frac{degree\_centrality(a)}{n-1}$$



Directed:
In/out-degree centrality

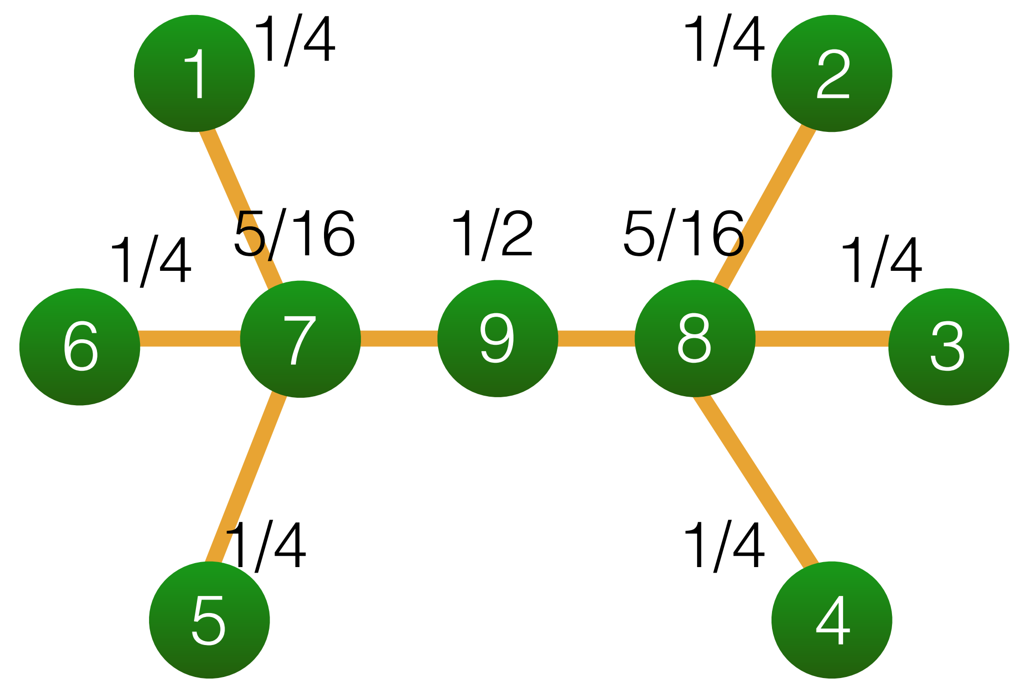


Eigenvector centrality

$$N = \{a_i, i = 1..n\},\ E = \{(a_j, b_j)\}$$

$$\text{centrality}(a) \sim \sum_{(a,b)\in E} \text{centrality}(b)$$

$$\lambda c = Ac \qquad \lambda_j,\ e_j$$



$$x \to Ax \to A^2x \to \dots A^j x$$

$$x = \sum x_i e_i \qquad x \to \sum_i (\lambda_i)^j x_i e_i \to e_1 + \sum_{i>1} \left(\frac{\lambda_i}{\lambda_1}\right)^j \frac{x_i}{x_1} e_i$$

$$\text{eigenvector\_centrality}(a) = e_1$$

Katz centrality

$N = \{a_i, i = 1..n\},\ E = \{(a_j, b_j)\}$ directed case

$$\text{centrality}(a) \sim \sum_{(b,a)\in E} \text{centrality}(b). \qquad \lambda_1 c = A^T c$$

$$\text{centrality}(a) = \alpha \sum_{(b,a)\in E} \text{centrality}(b) + \beta$$

$$\alpha < 1/\lambda_1$$

Pagerank

$$N=\{a_i, i=1..n\},\ E=\{(a_j,b_j)\}$$

$$\text{centrality}(a) \sim \sum_{(a,b)\in E} \frac{\text{centrality}(b)}{degree(b)}$$

$$\alpha, 1-\alpha \qquad \alpha = 0.85$$

$$\text{pagerank}(a) = \alpha \sum_{(a,b)\in E} \frac{\text{pagerank}(b)}{degree(b)} + \frac{1-\alpha}{n}$$

$$\text{pagerank}(a) = \alpha \sum_{(b,a)\in E} \frac{\text{pagerank}(b)}{out\_degree(b)} + \frac{1-\alpha}{n}$$