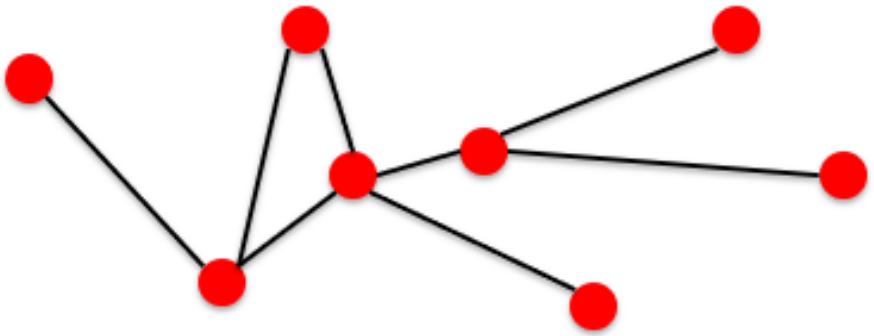
Network Analysis and Visualization

Veera Muangsin

Network represents relationships



- **components**: nodes, vertices — N
- **interactions**: links, edges — L
- **system**: network, graph — (N,L)

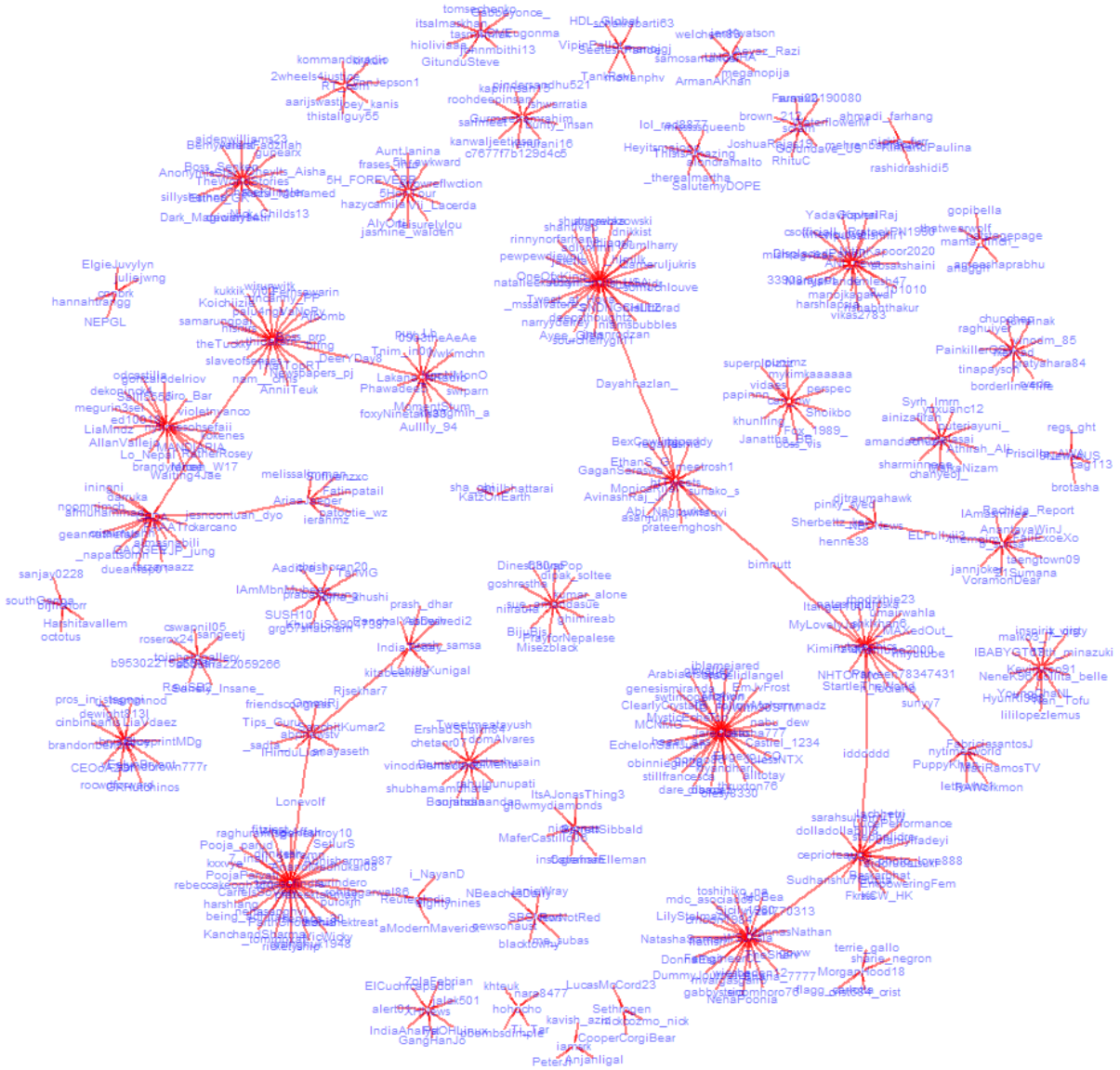
Examples

- Contacts (telephone, email)
- Social networks (friend, follow, share)
- Reference (citation, hyperlinks)
- Collocation/co-occurrence (words, events)
- Connections (roads, computer network)

Applications

- Social Network Analysis
- Bibliographic Analysis
- Transportation/communication

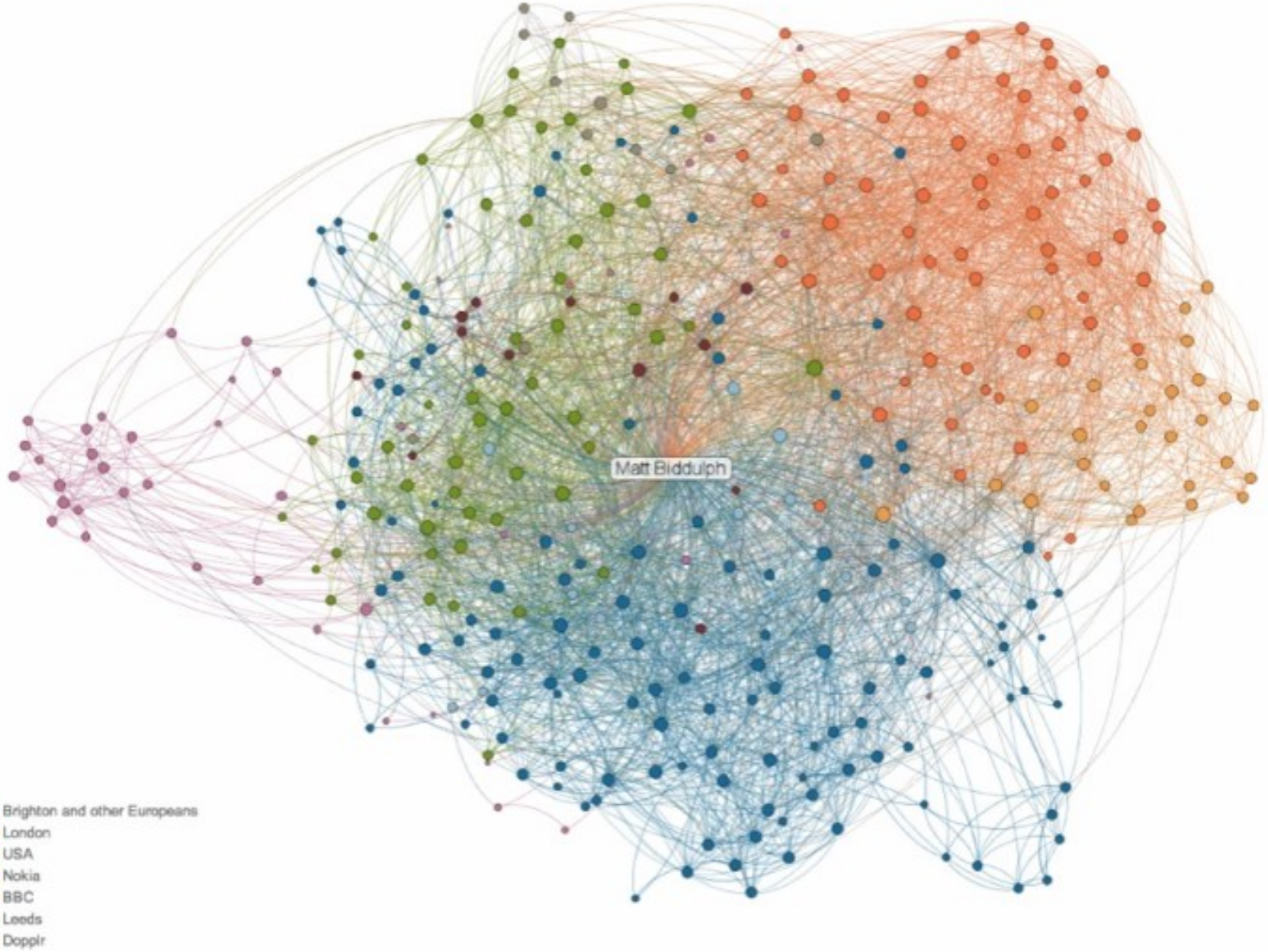
Retweet Network



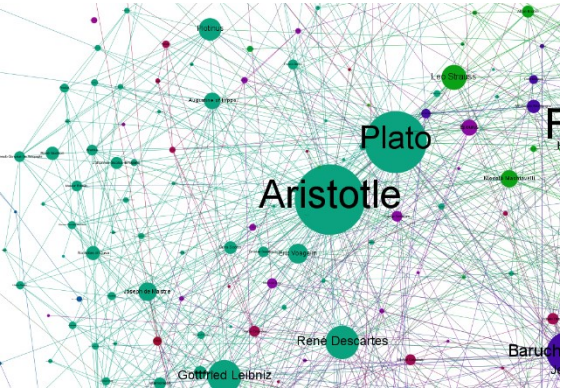
LinkedIn Maps

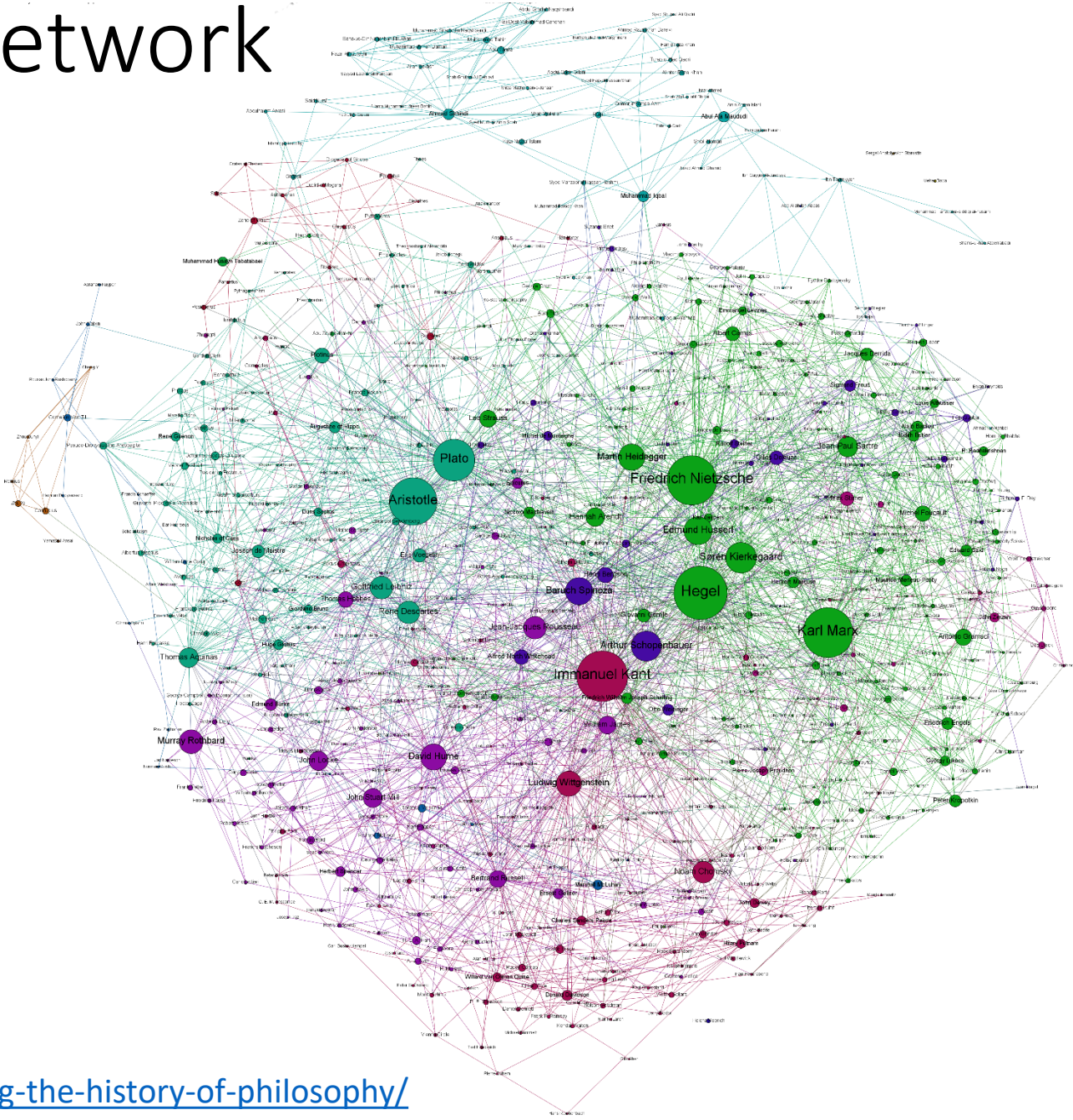
Matt Biddulph's Professional Network

as of January 24, 2011



Philosophy Network





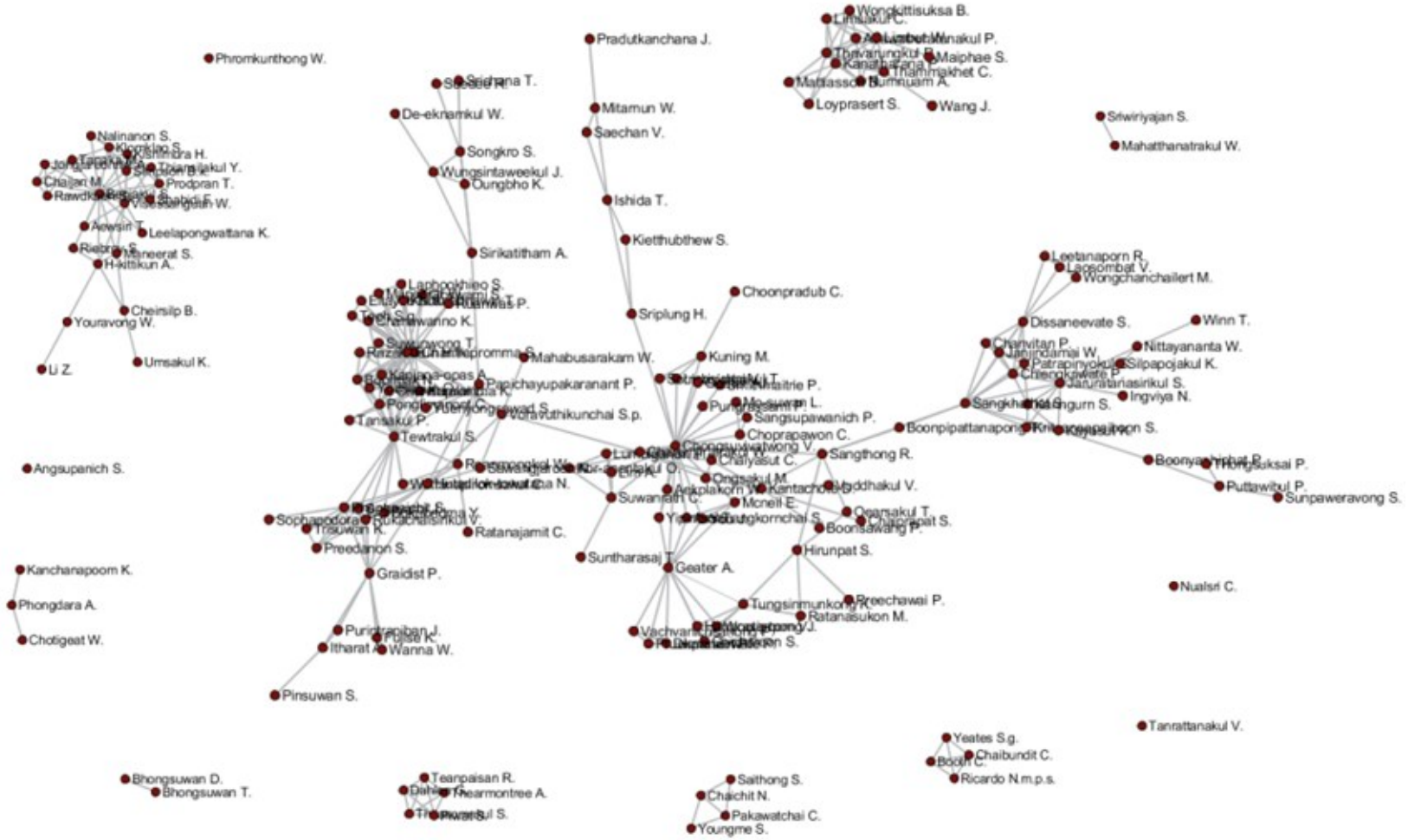
http://coppelia.io/2012/06/graphing-the-history-of-philosophy/

Facebook Friendships (2010)

Each line connects between two cities, weighted by the number of friendships and geographic distance.

https://www.facebook.com/notes/10158791468612200/
https://paulbutler.org/2010/visualizing-facebook-friends/

Co-authorship Network



Ingredient Complement Network (word collocation)

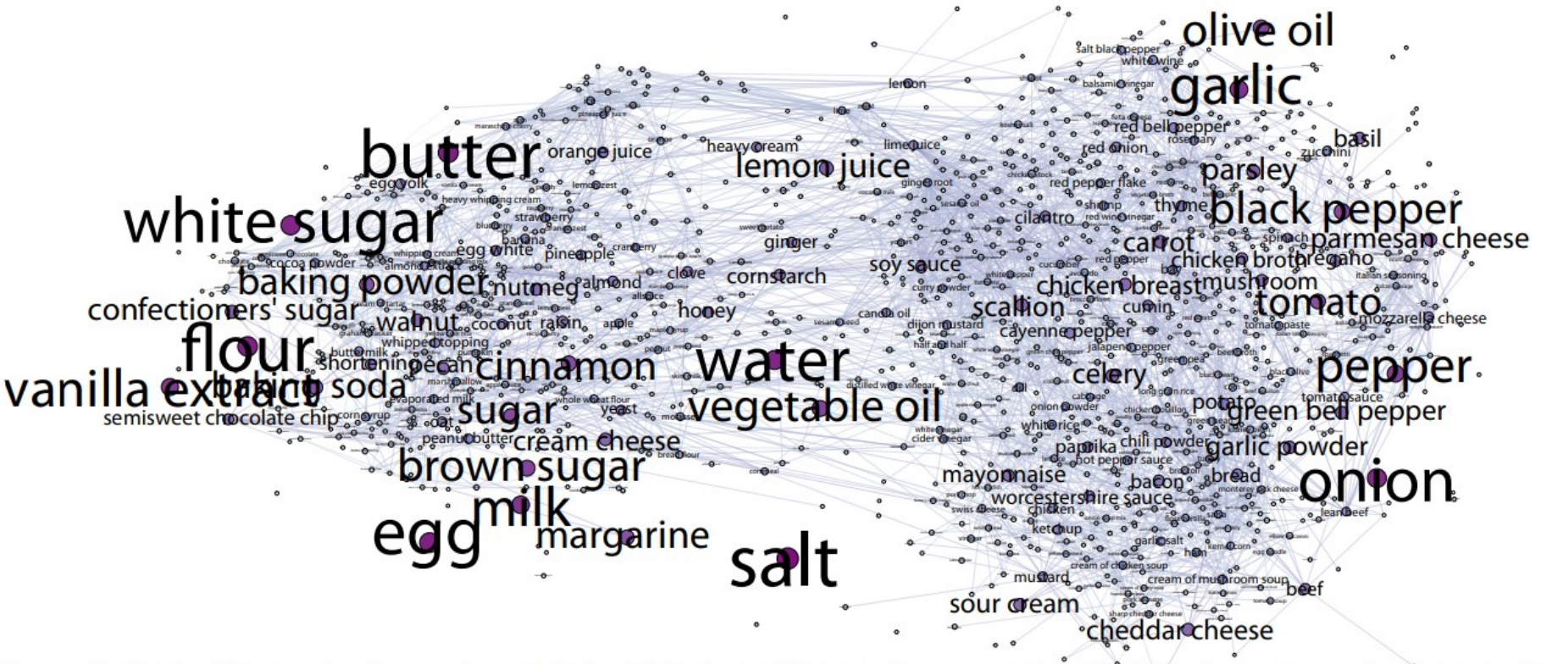
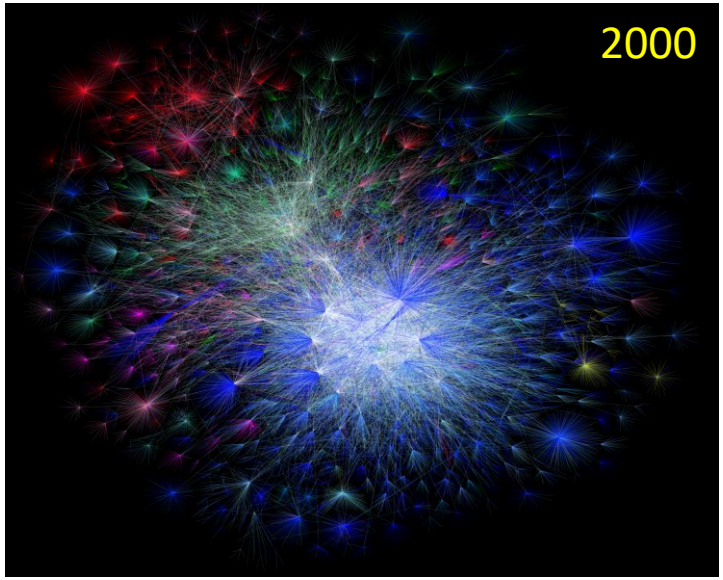
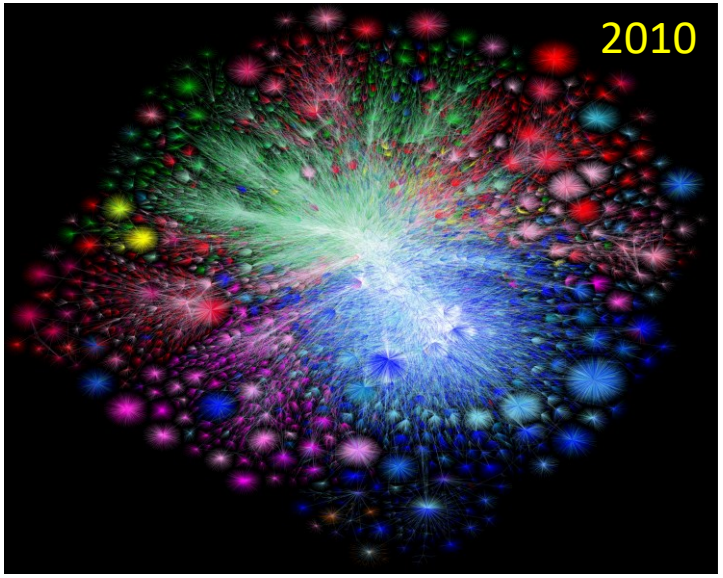


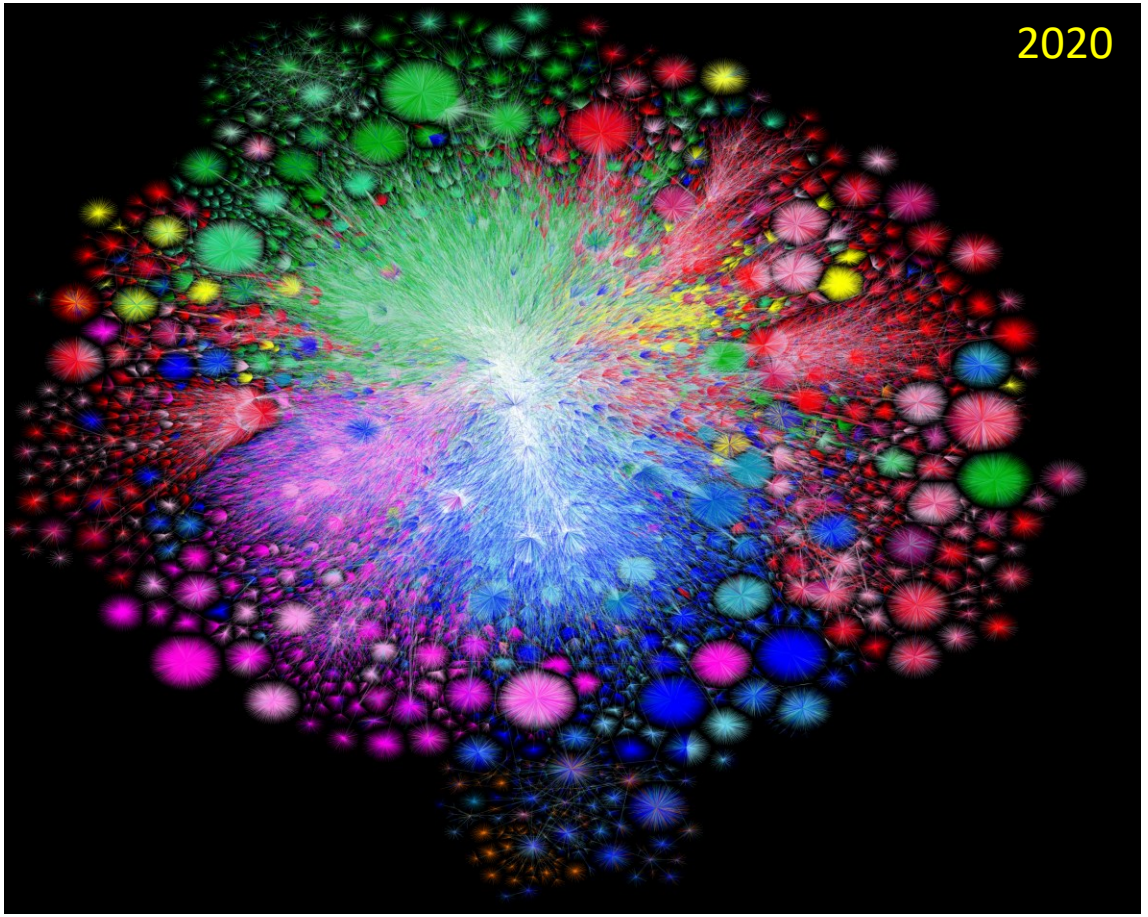
Figure 2: Ingredient complement network. Two ingredients share an edge if they occur together more than would be expected by chance and if their pointwise mutual information exceeds a threshold.

"Recipe recommendation using ingredient networks" Chun-Yuen Teng, Yu-Ru Lin, Lada A. Adamic
Proceedings of the 4th Annual ACM Web Science Conference 2012 (WebSci '12)

Map of the Internet

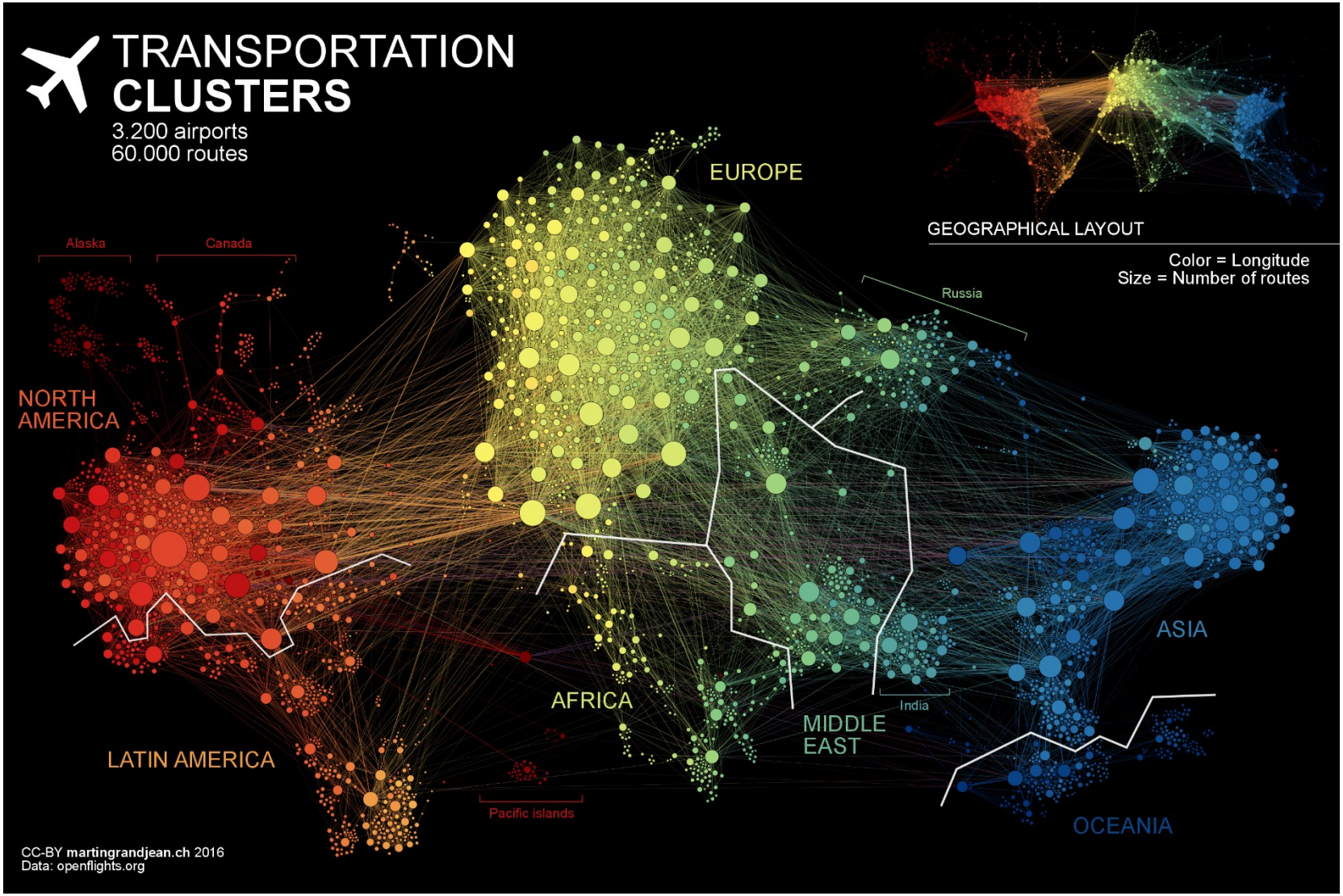






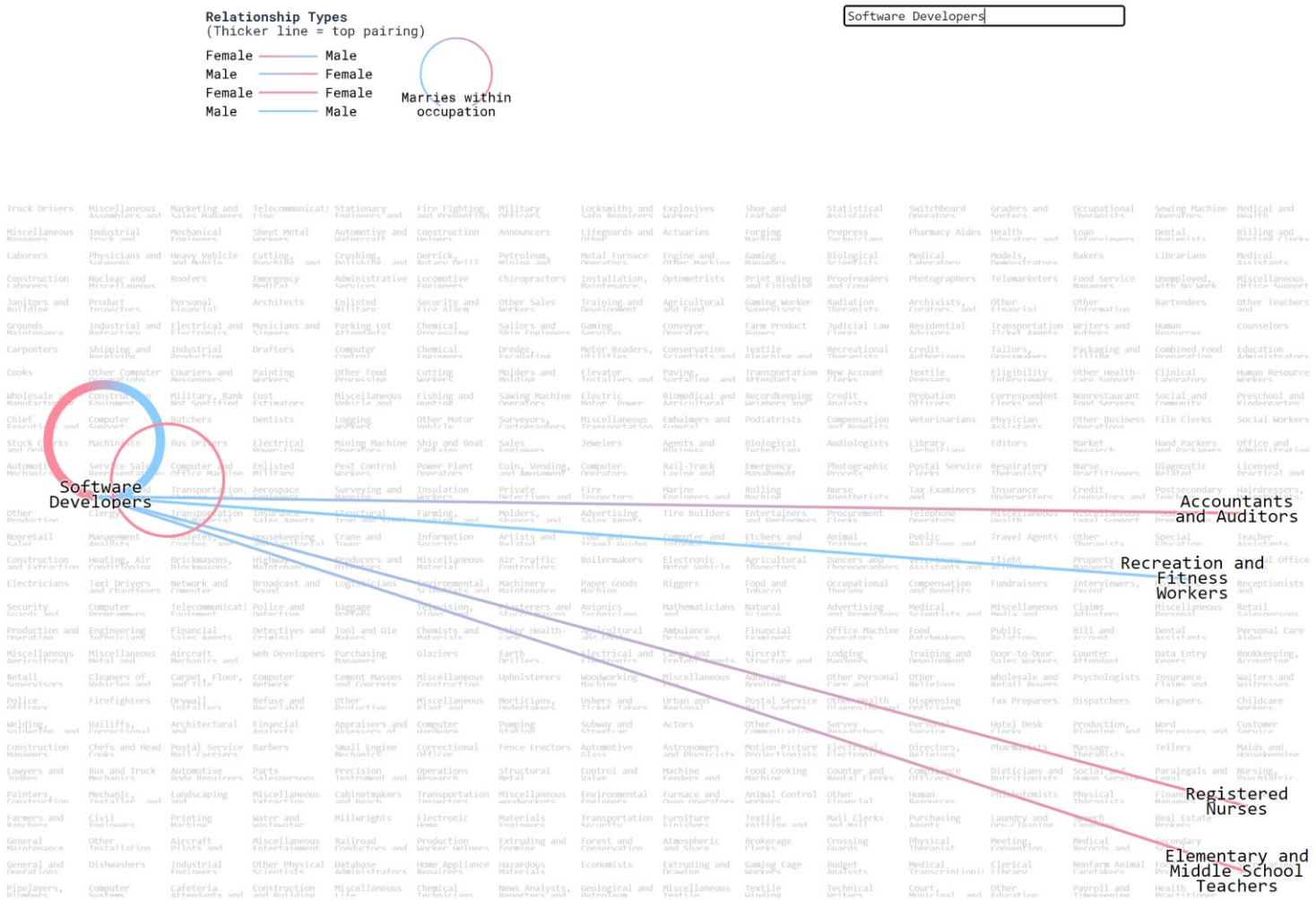
https://www.opte.org/the-internet
http://renderbot.nyc1.opte.org/

Air Traffic Network



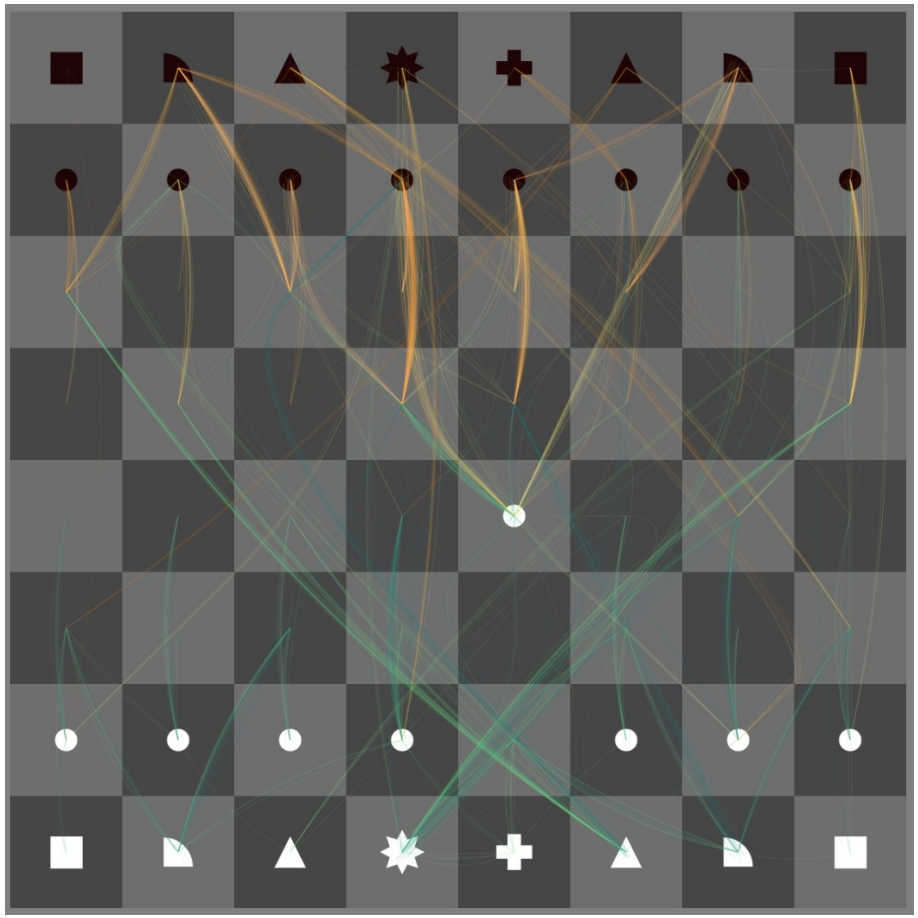
http://www.martingrandjean.ch/connected-world-air-traffic-network/

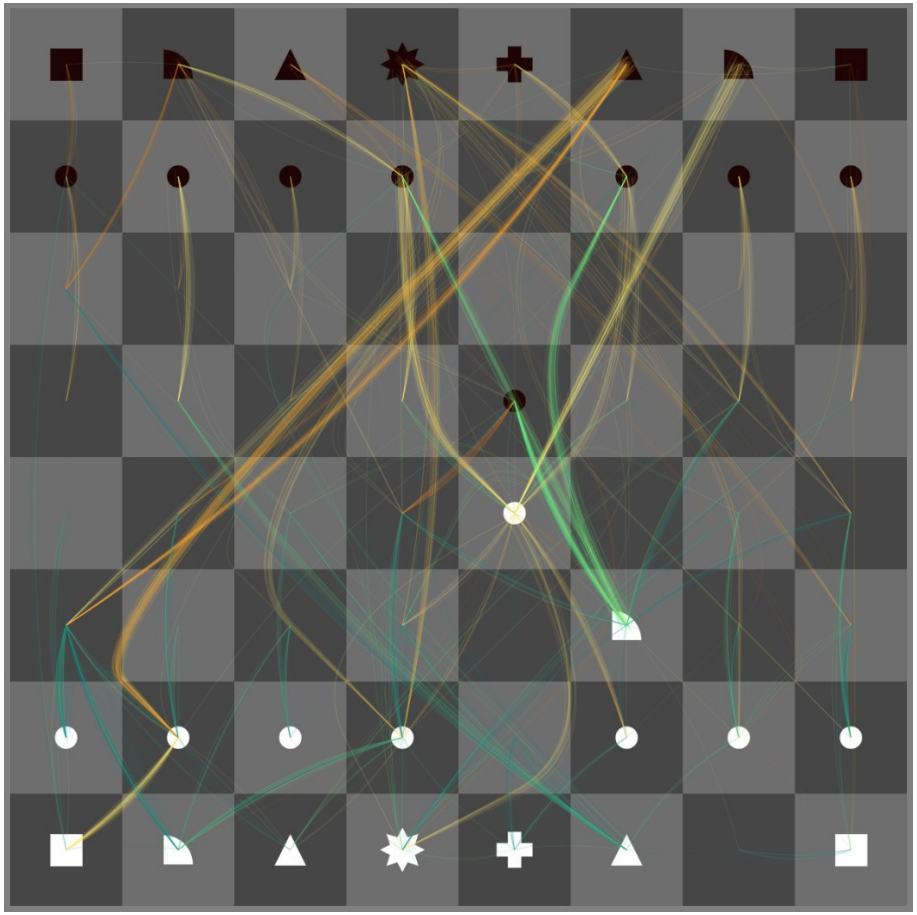
Who marries whom (interactive)



https://www.bloomberg.com/graphics/2016-who-marries-whom/

Thinking Machine (interactive)





https://www.bewitched.com/chess/

Network Models

Basic network models

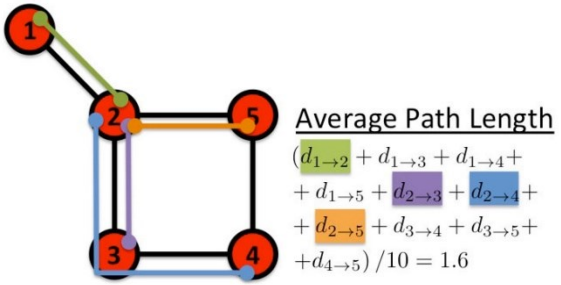
- Random network
- Small-world network
- Scale-free network

Key Network Features

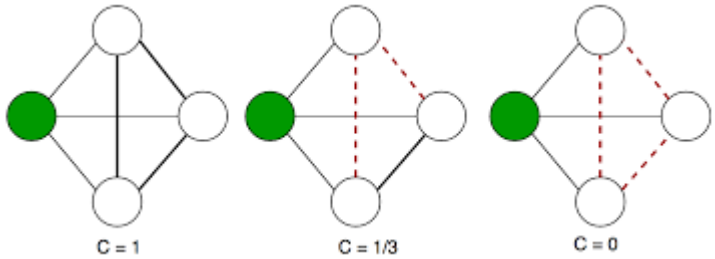
- Degree Distribution
  - Distribution function of node degrees
  - P(k) = fraction of nodes with degree k

$P = fraction\ of\ nodes\ with\ degree\ k$

- Average Path Length
  - Average number of hops between any two nodes



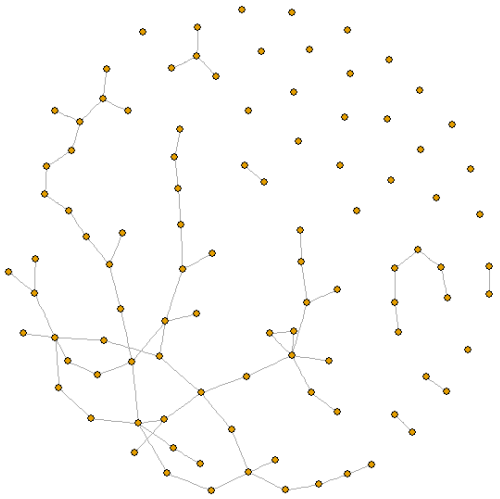
- Clustering coefficient
  - Average fraction of pairs of neighbors of a node that are also neighbors of each other.

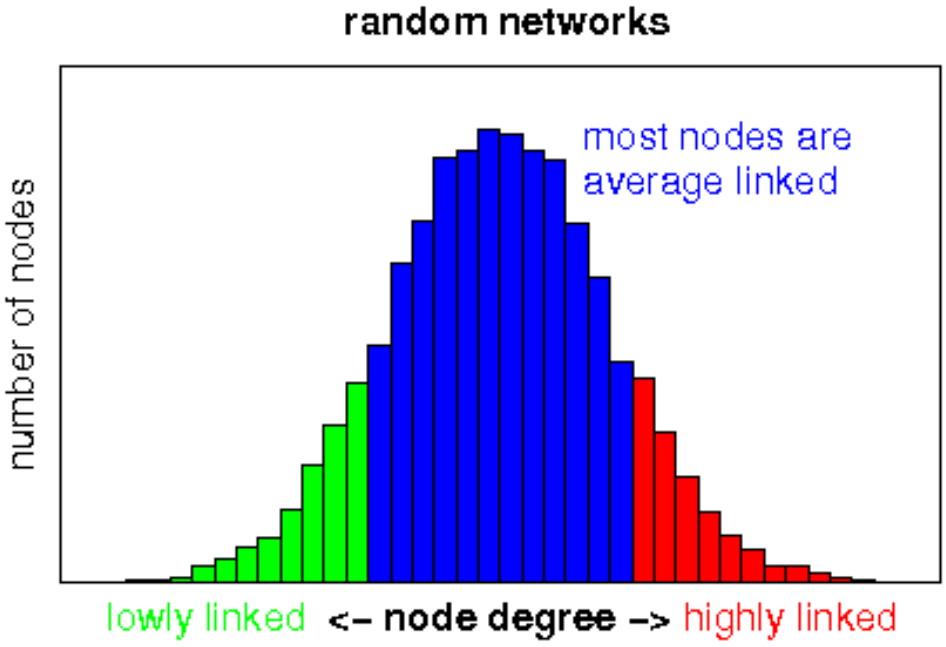


$$C = \frac{no.\ of\ triangles}{no.\ of\ all\ possible\ triangles}$$

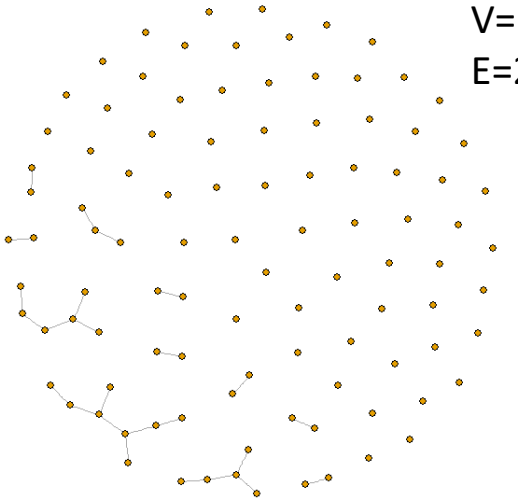
Random Network

- Known as Erdös-Rényi graph or binomial graph
- A random network is obtained by starting with a set of *n* isolated vertices and adding successive edges between them at random.
- A giant component (largest connected component) will emerge.
- Degree distribution is normal distribution.
- Random networks may not represent real-world networks but offer base-line comparison to other models.

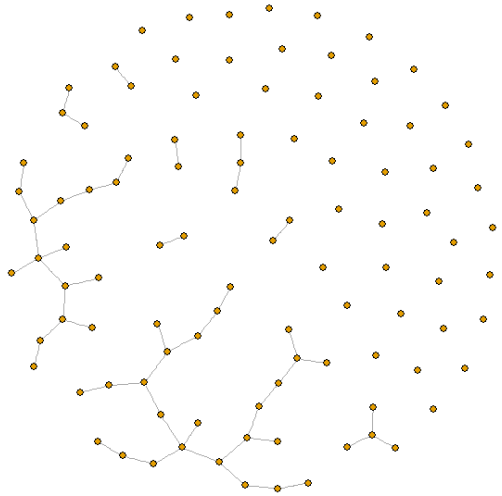




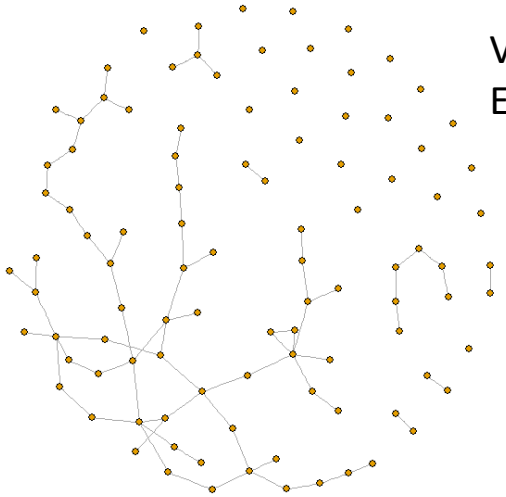
Random network



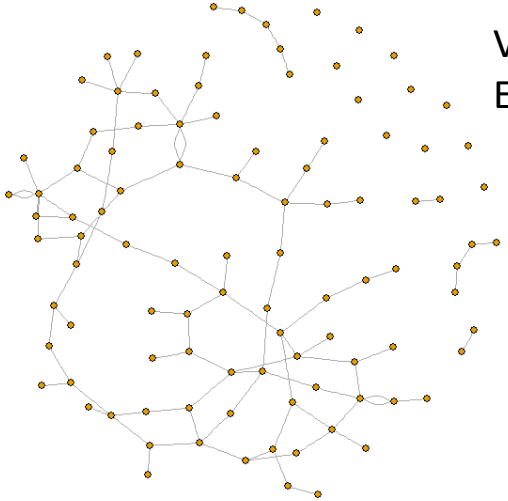
V=100
E=25



V=100
E=50



V=100
E=75



V=100
E=100

Small-world Network

- Known as Watts-Strogatz graph
- Network with high clustering coefficient and short average path length
- Most nodes are connected to a few nearby neighbors but some nodes also have connections to far away nodes.
- That shortens the average path length and therefore most nodes can be reached from every other node by a small number of hops.
- Real-world example of small-world network
  - Social connection
  - Email, telephone call-graph

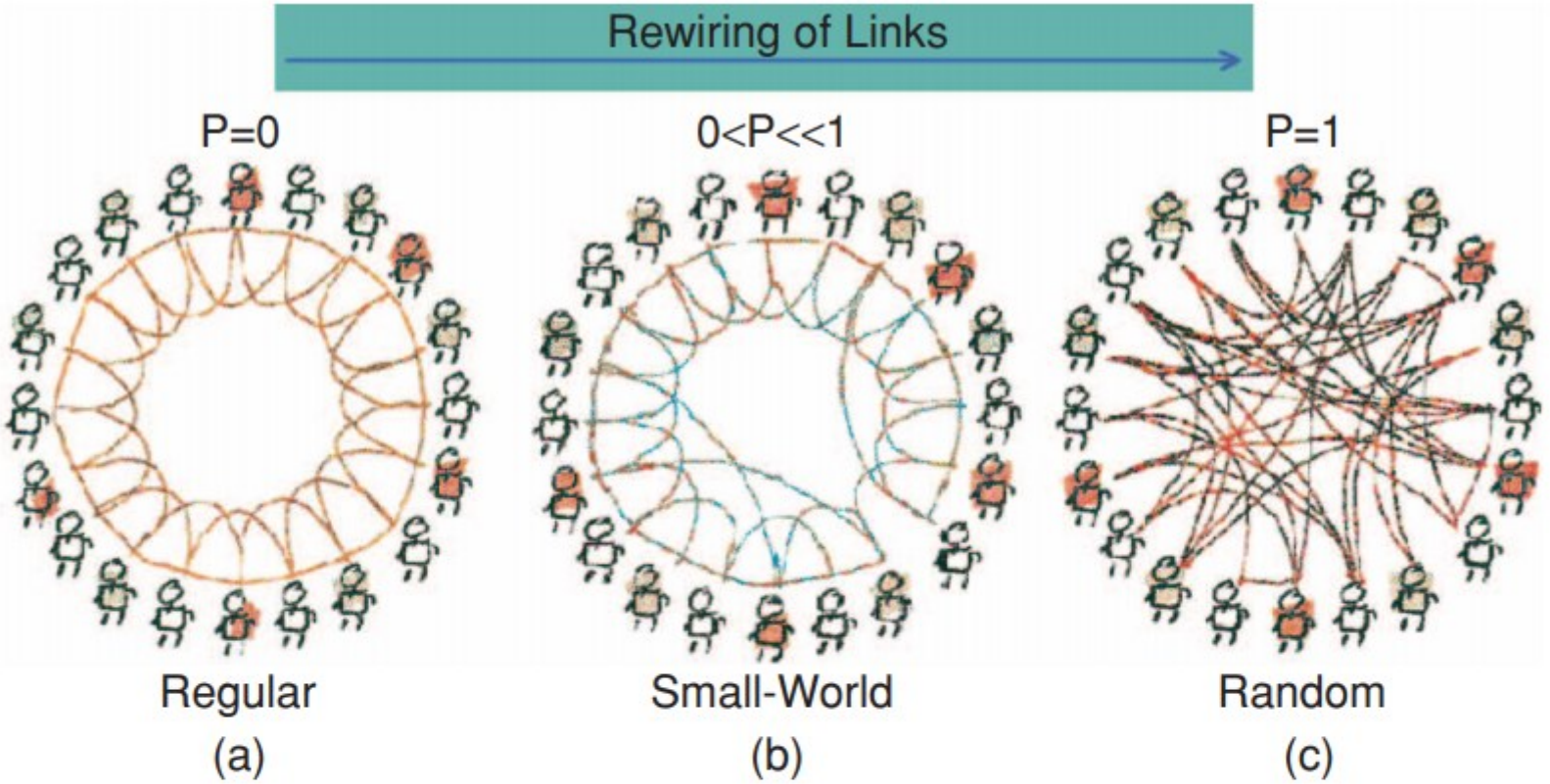
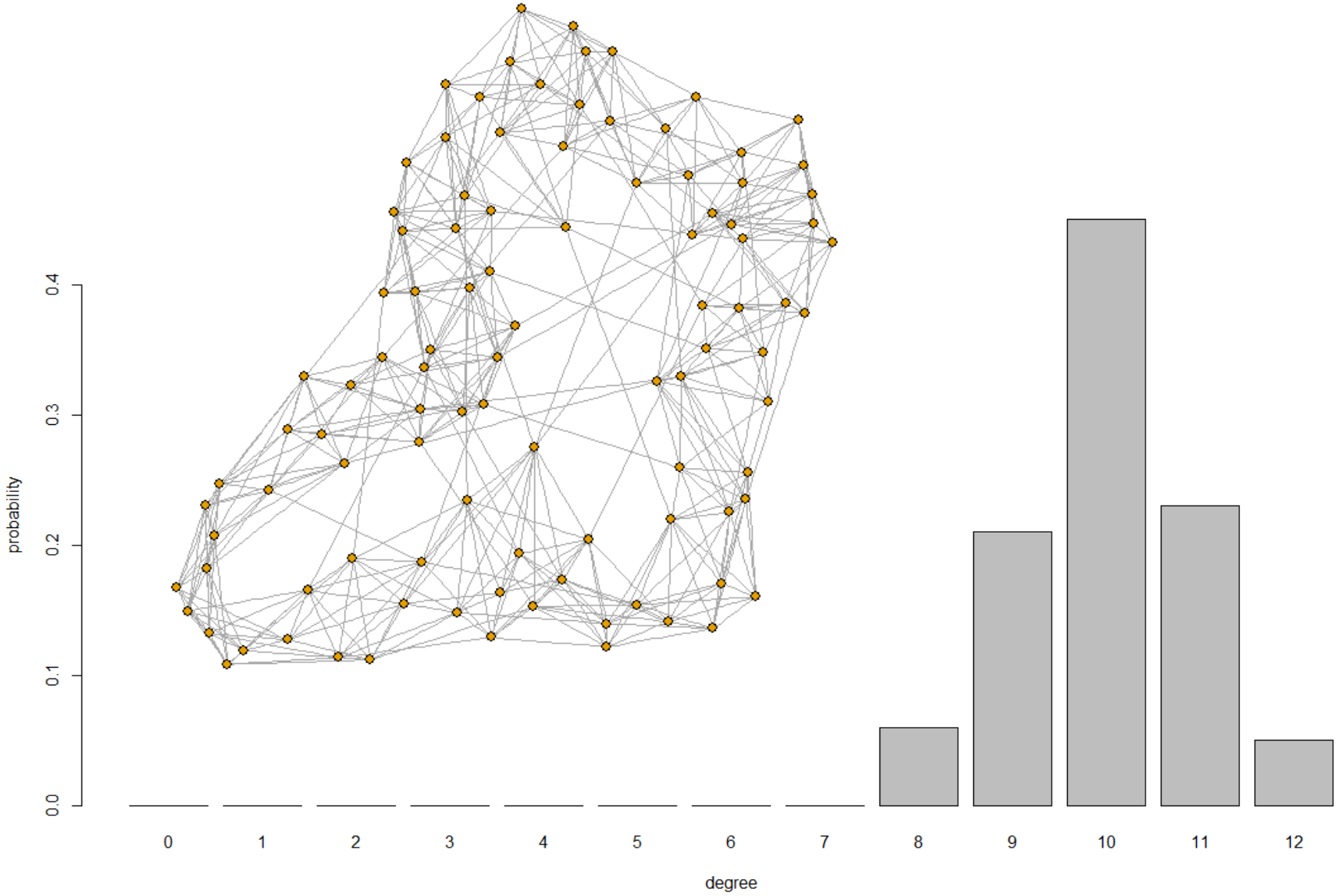


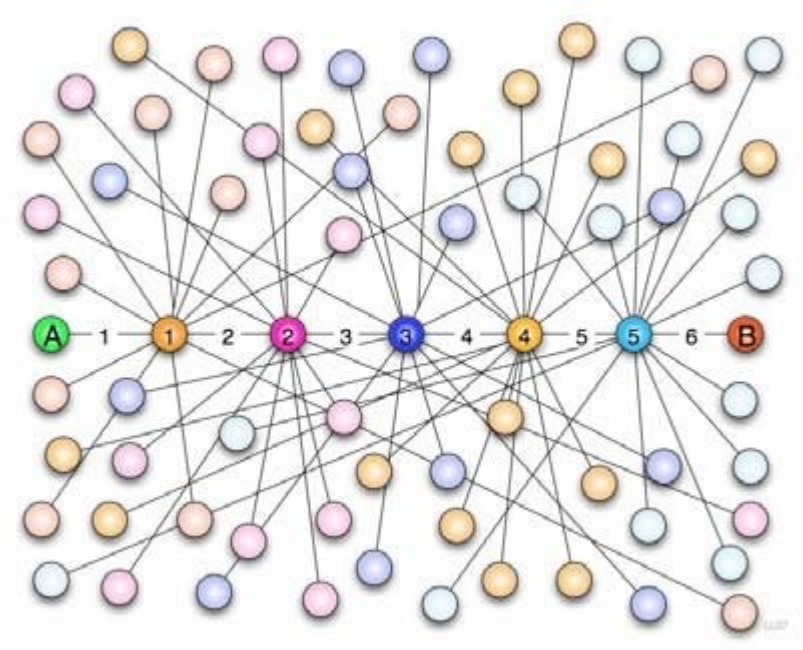
Figure 7. (a) In this completely regular friendship network, people are friends with only their 4 nearest neighbors. The network is highly cliquish, and any 2 people are on average many degrees apart. (b) In this small-world network, people still know 4 others on average, but a few have distant friends. The network is still highly cliquish, but the average degree of separation is small. (c) In this random network, everyone still knows 4 others on average, but friends are scattered: few people have many friends in common, and pairs are on average only a few degrees apart.

A Small-World Network



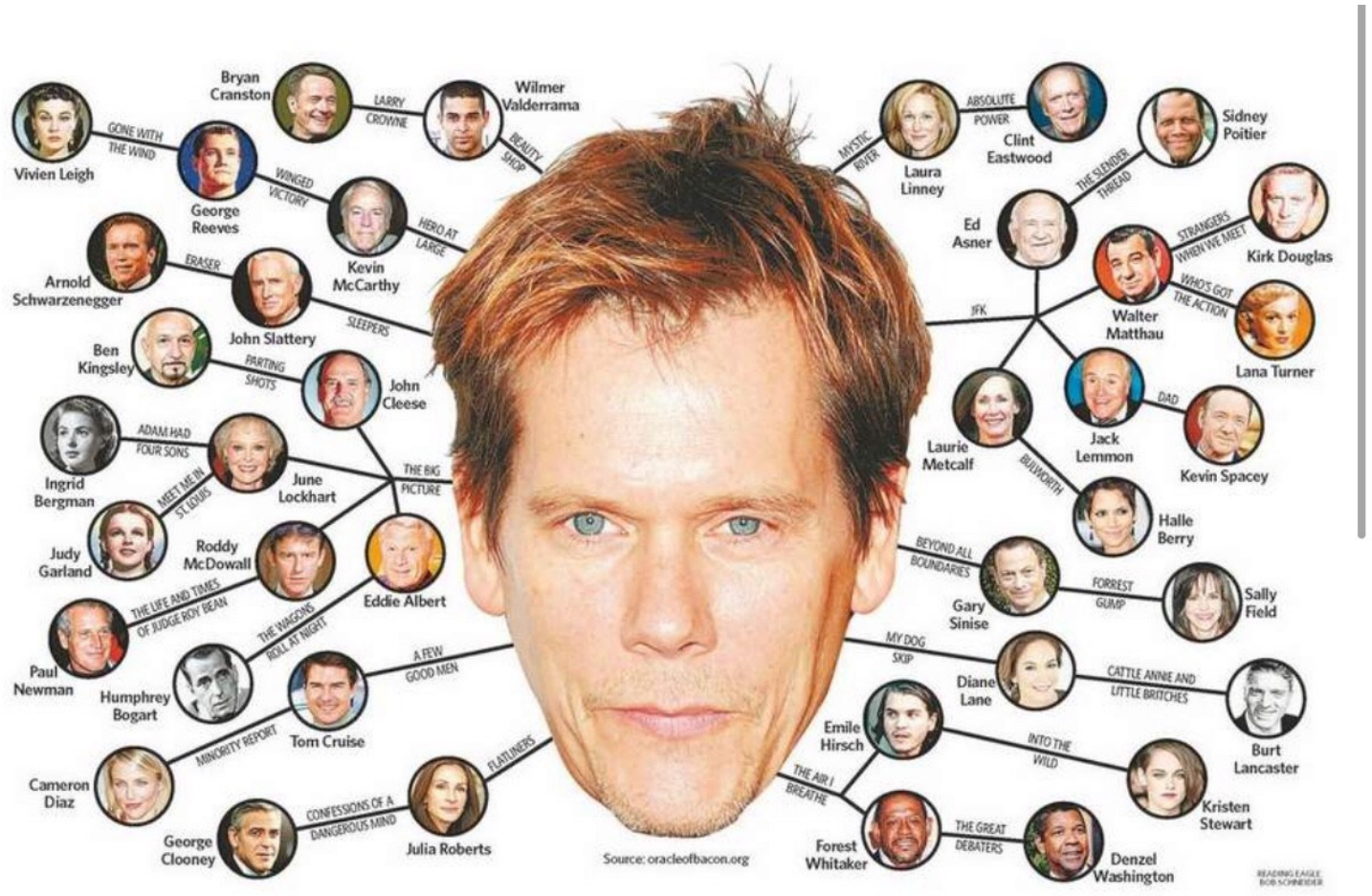
Six degrees of separation

- A theory that anyone can connect to anyone else in the world via at most six steps of social connection.



Six degree of Kevin Bacon

- A prediction that any actor can be linked through their film roles to Bacon within six steps



https://www.youtube.com/watch?v=n9u-TITxwoM

SIX DEGREES OF WIKIPEDIA

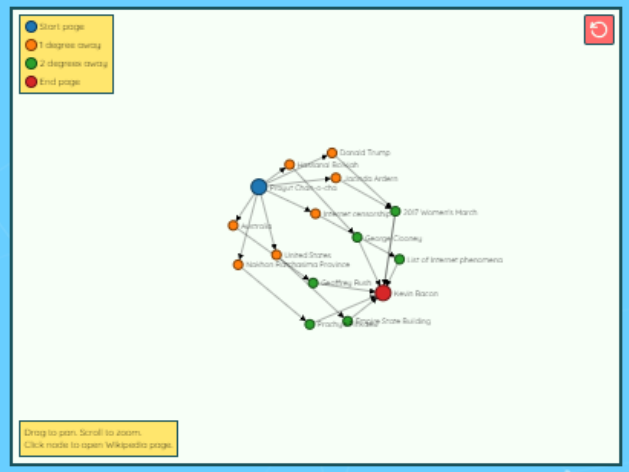
Find the shortest paths from

Prayut Chan-o-cha ⇄ Kevin Bacon

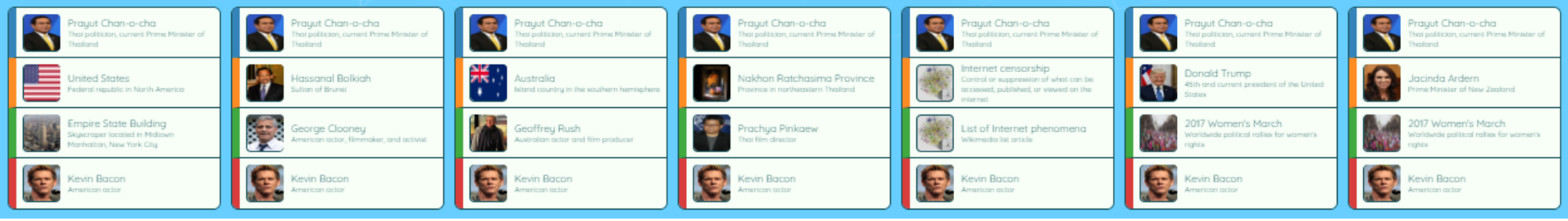
Go!

Found **7 paths** with **3 degrees** of separation from **Prayut Chan-o-cha** to **Kevin Bacon** in **2.54 seconds!**

Tweet result



Individual paths

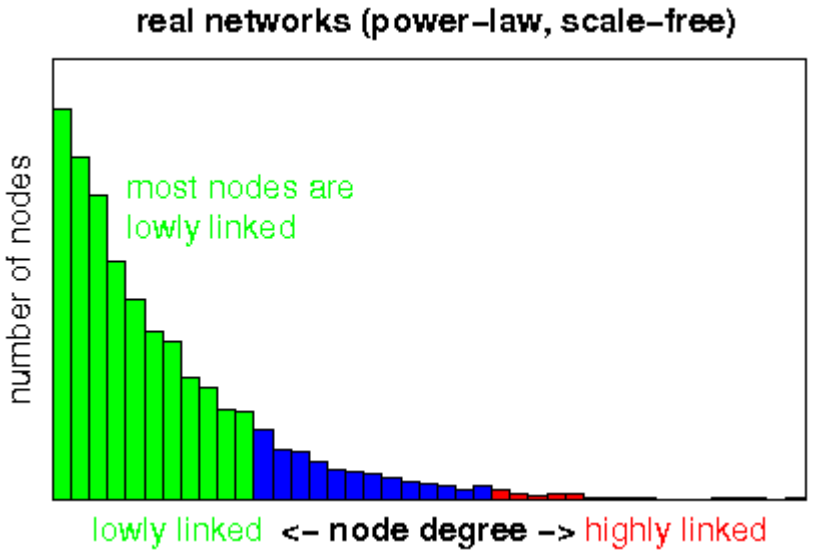


https://www.sixdegreesofwikipedia.com/?source=Prayut%20Chan-o-cha&target=Kevin%20Bacon

Scale-Free Network

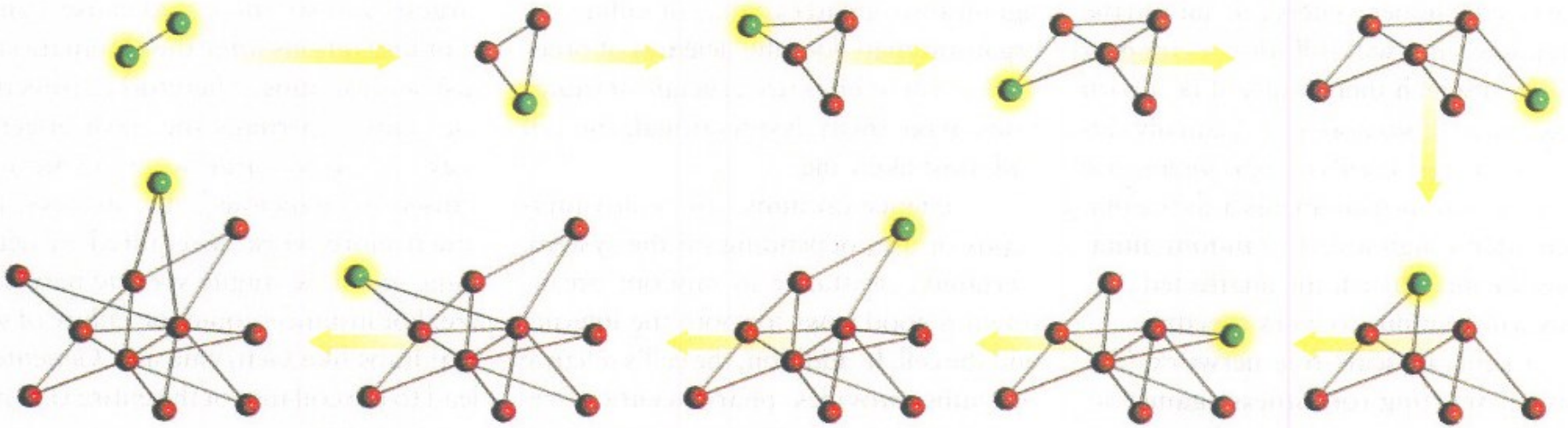
- Known as Barabási-Albert graph
- This kind of networks grow through two key principles
  - Growth: network start small and grows over times as new nodes join
  - Preferential attachment
    - New nodes tend to connect to already-popular nodes
- Degree distribution follows a power law
  - Its graph has a long tail or hockey stick shape
- Many real-world networks are thought to be scale-free
  - World wide web
  - Social media network
  - Citation network
  - Airline route network

It's call "scale-free" because:
- There is no 'typical' or 'average' number of connections that is considered 'normal'.
- The pattern repeats at every scale (sampling any number of nodes and you'll see the same pattern)

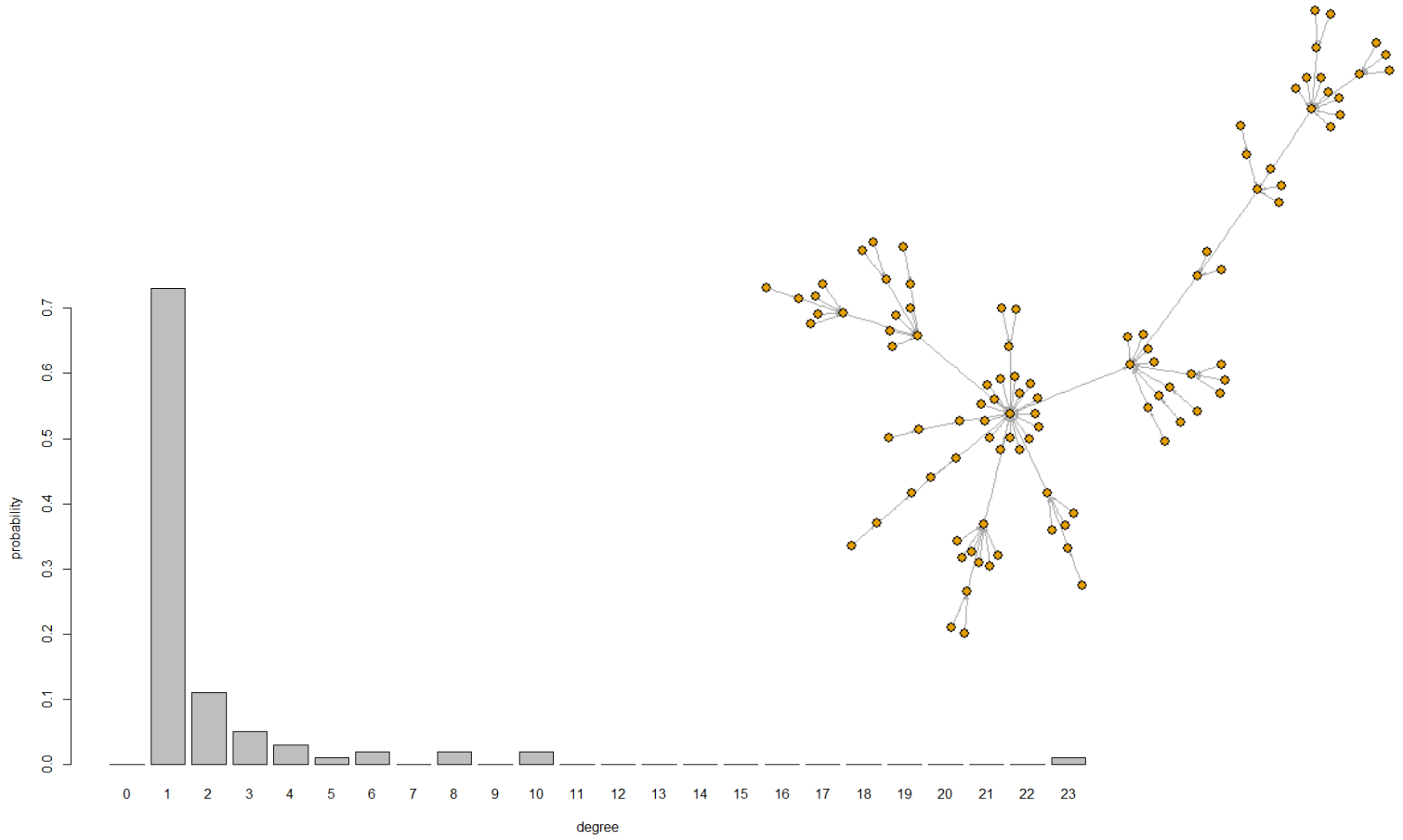


BIRTH OF A SCALE-FREE NETWORK

A SCALE-FREE NETWORK grows incrementally from two to 11 nodes in this example. When deciding where to establish a link, a new node (*green*) prefers to attach to an existing node (*red*) that already has many other connections. These two basic mechanisms—growth and preferential attachment—will eventually lead to the system's being dominated by hubs, nodes having an enormous number of links.



A Scale-Free Network



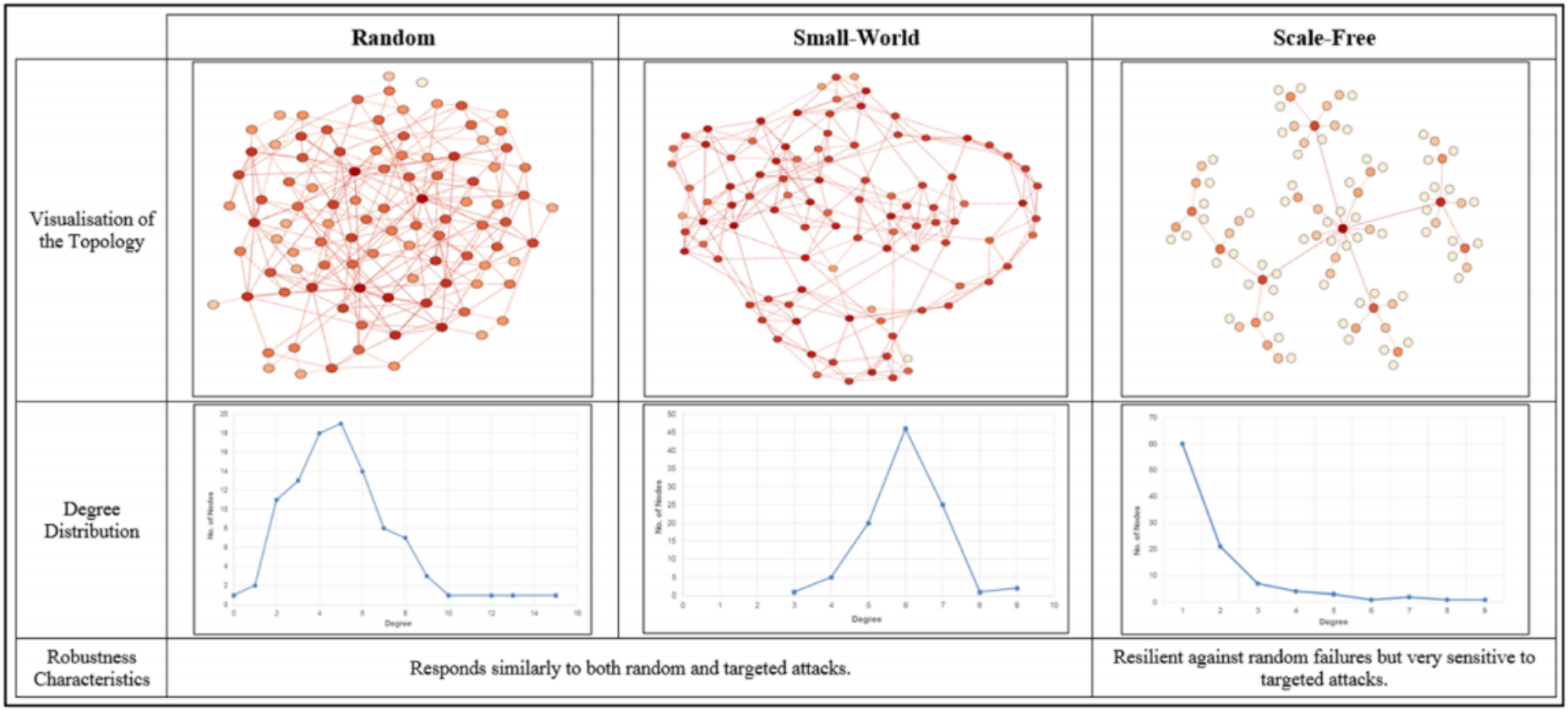


Fig. 1 Comparison of random, small-world and scale free networks. Topological structure of benchmark network models. Random and Small-world network topologies do not include hub nodes. In contrast, scale-free topologies are characterised by the presence of small number of highly connected hub nodes and a high number of feebly connected nodes. Presence of distinct hubs in scale-free networks make them more vulnerable to targeted attacks, compared to random and small-world networks

Centrality Measures

Network Centrality

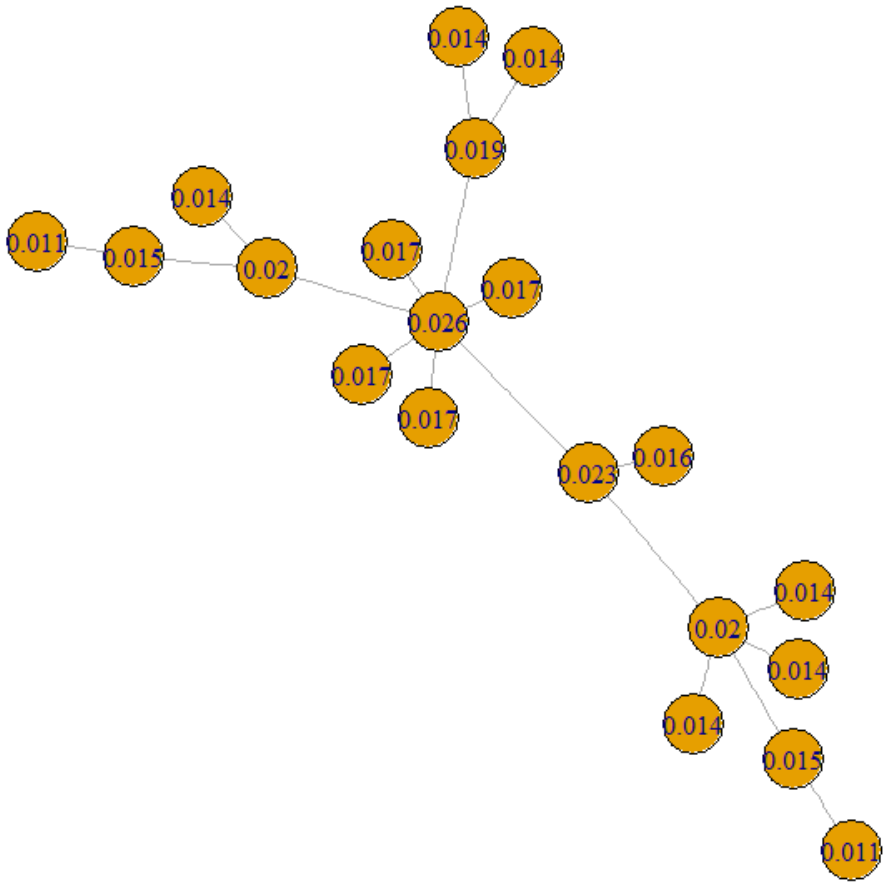
- Which nodes in the network are most important? Or most central?
- How centrality distributed among nodes?
- Applications
  - Identify influencer in social network
  - If I am to pass on a message to three people in this network so that they in turn convey it to their friends and so on. Which three people should I select?
  - If I were to nominate a leader for a team of 500, whom should I pick?

Centrality Measures

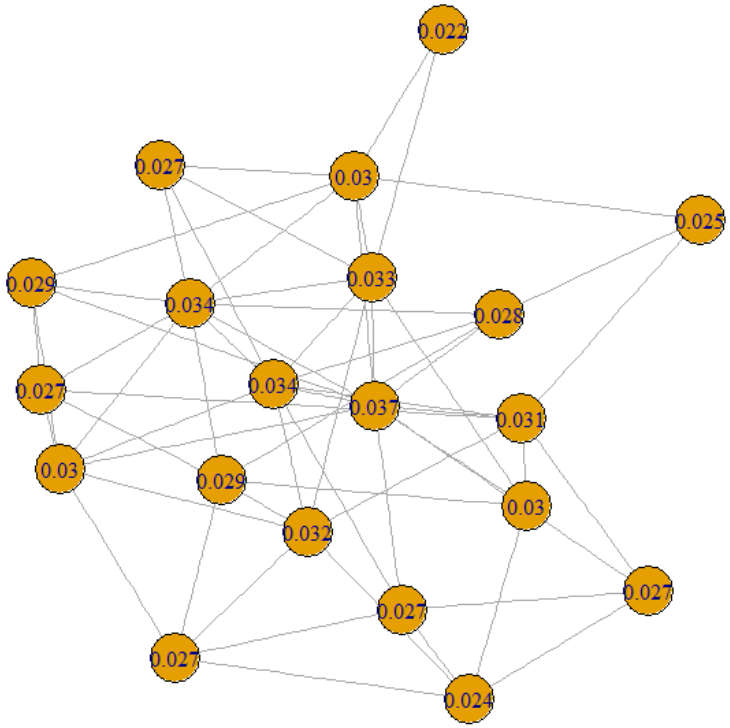
- Degree centrality
  - An important node is involved in large number of interactions.
- Closeness centrality
  - An important node can communicate quickly with other nodes.
- Betweenness centrality
  - An important node will lie on most of the shortest paths between other nodes.
- Eigenvector centrality
  - An important node is connected to important neighbors.

https://sites.google.com/site/networkanalysisacourse/schedule/an-introduction-to-centrality-measures

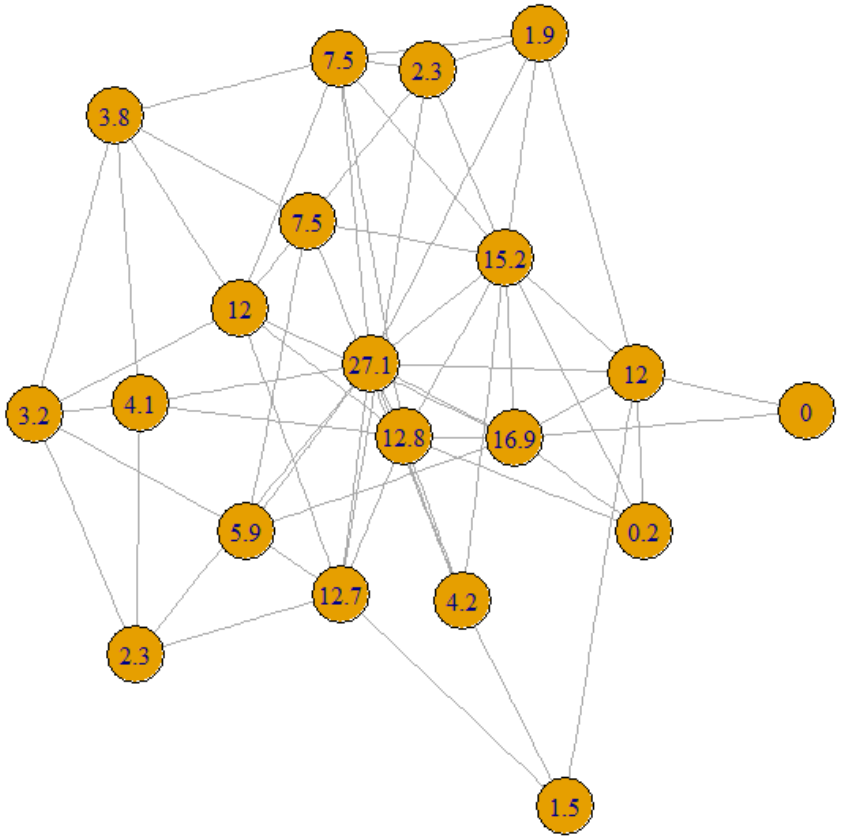
Closeness Centrality: scale-free network



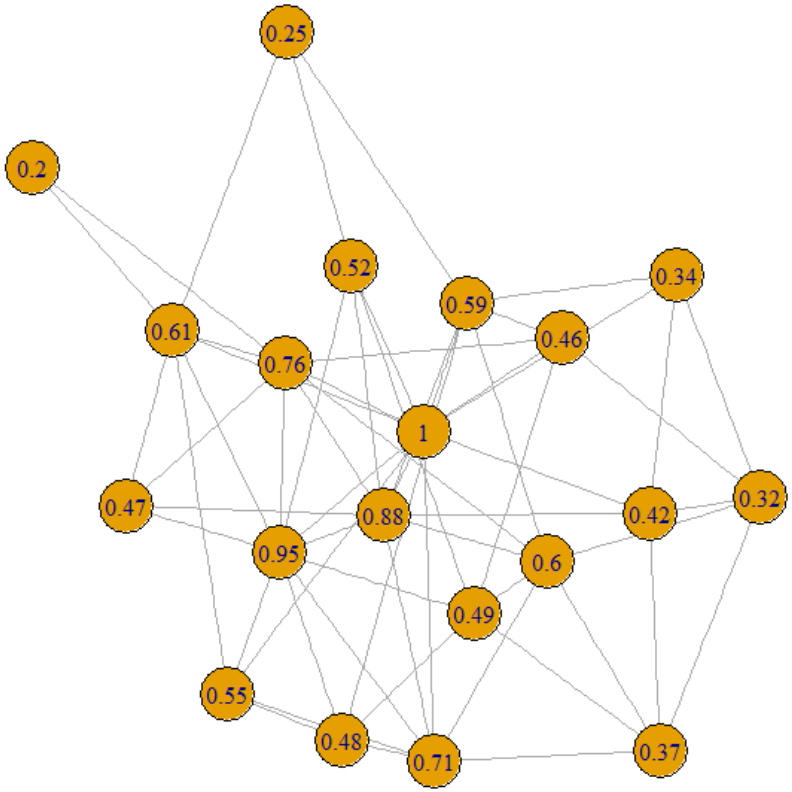
Closeness Centrality: small-world network



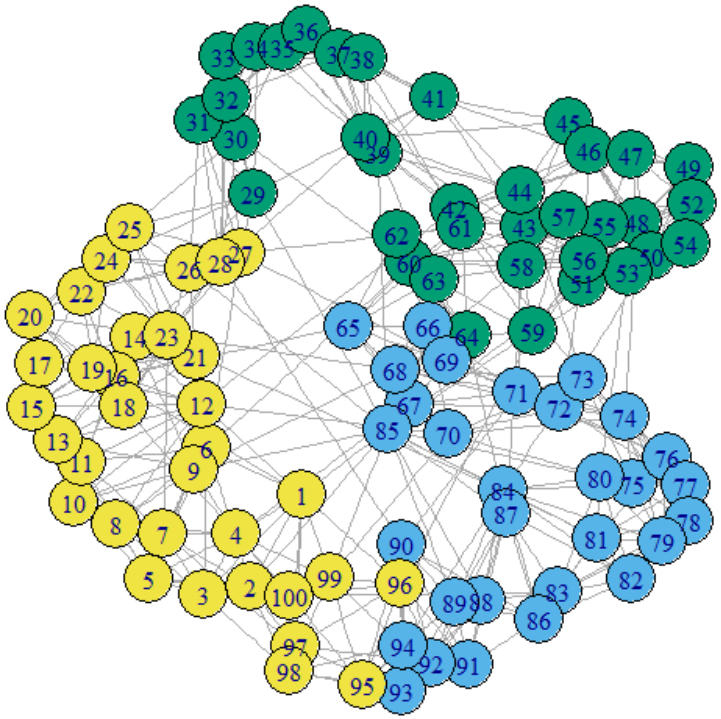
Betweenness Centrality: small-world network



Eigenvector Centrality : small-world network



Community Detection



Network Visualization

Network Visualization Styles

Node–Link Diagrams

Connection Marks

✔ NETWORKS ✔ TREES

Adjacency Matrix

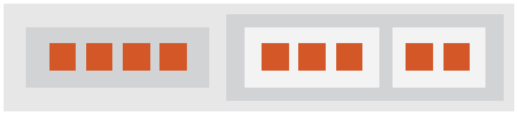
Derived Table

✔ NETWORKS ✔ TREES

Enclosure

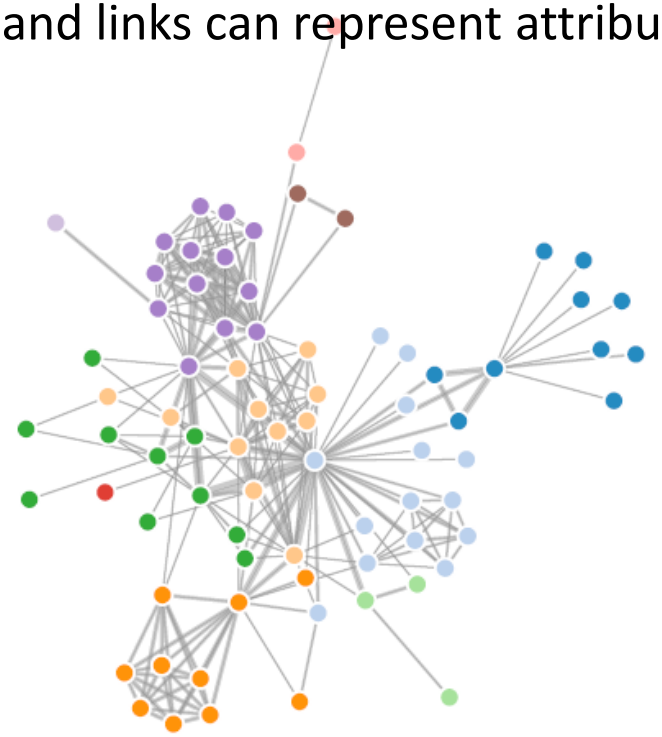
Containment Marks

✘ NETWORKS ✔ TREES



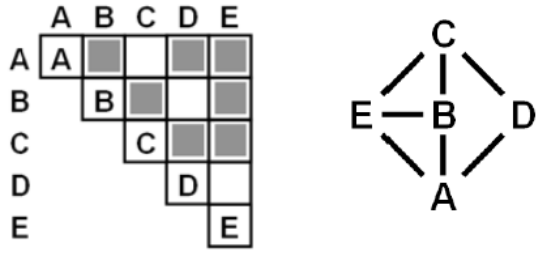
Node-Link Diagrams

- Link represents relation
- Good for exploring topology, locating clusters, following path
- Cluttering with high edge:node ratio
- Layout matters
- Size and color of nodes and links can represent attributes

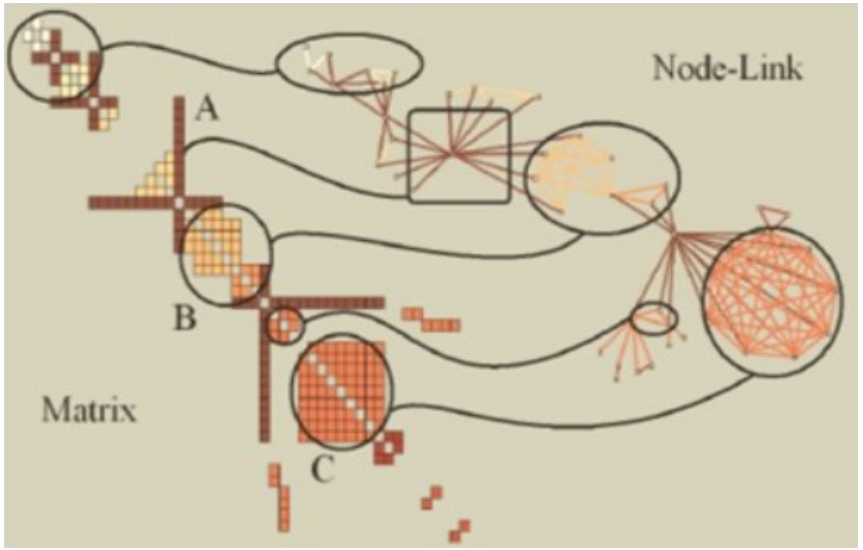


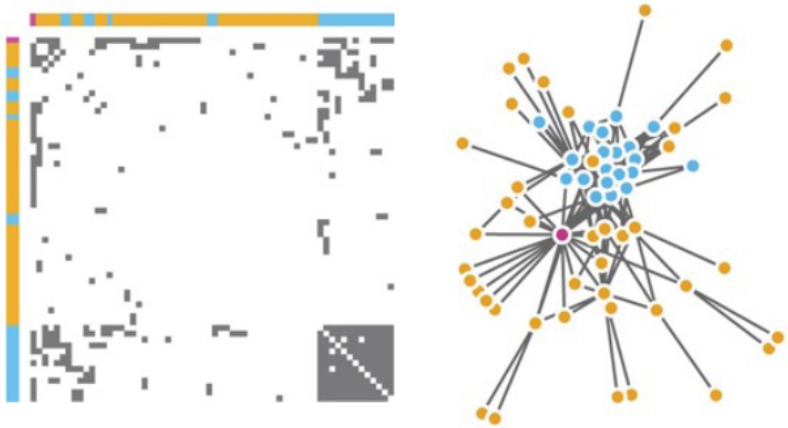
Adjacency Matrix

- Each marked cell represents an edge.
- Long straight lines represent high-degree nodes
- Hot spots represent clusters
- Good for dense network.
- Hard to follow paths.
- Node ordering matters.



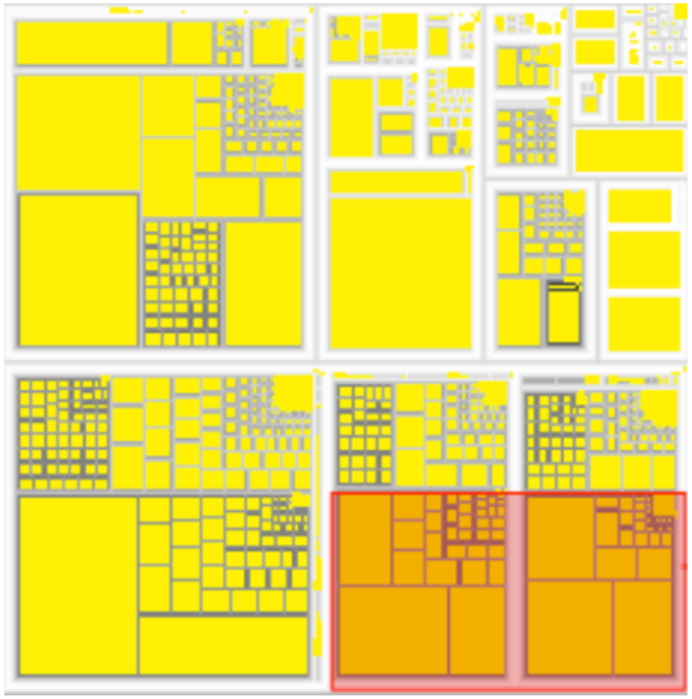
[NodeTrix: a Hybrid Visualization of Social Networks. Henry, Fekete, and McGuffin. IEEE TVCG (Proc. InfoVis) 13(6):1302-1309, 2007.]





Enclosure

- TreeMap
- For tree only
- Sub-areas represent children nodes



Force-Directed Layout

- Spring model
  - Edges – spring (gravity, attraction)
  - Nodes – charged particles (repulsion)
- Iteratively recalculate positions of nodes according to forces on each node until the network is stable.

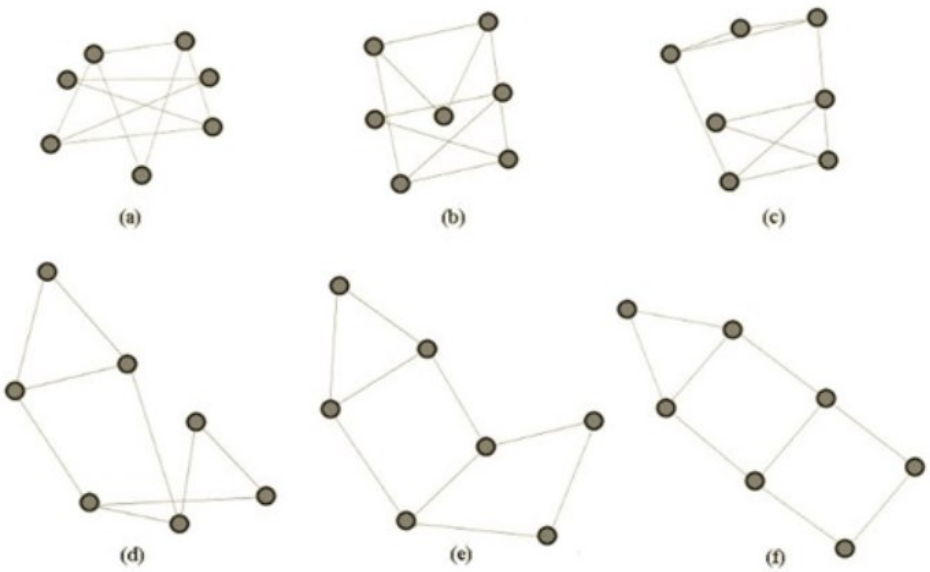
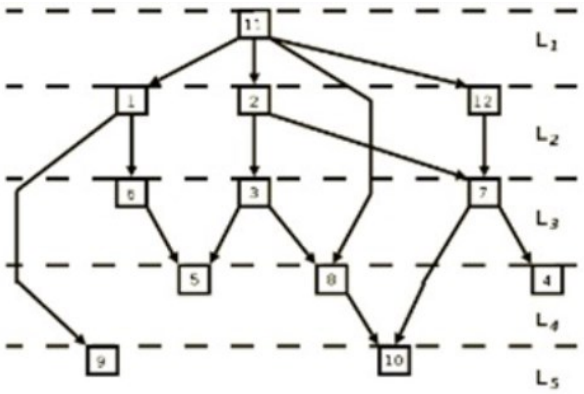
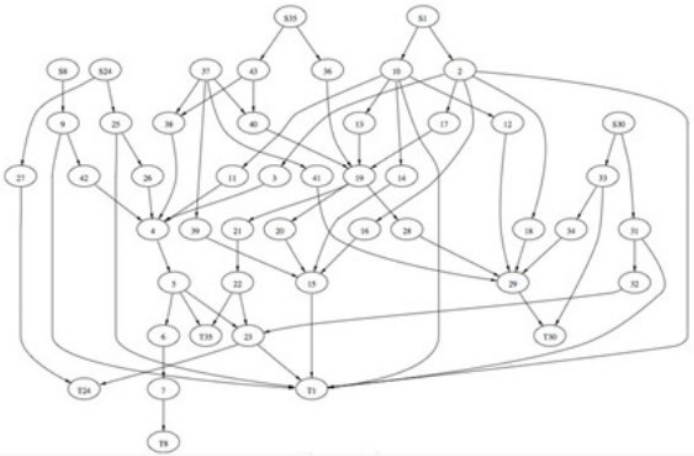


Figure 2: A graph drawing through a number of iterations of a force directed algorithm.

Hierarchical Layout

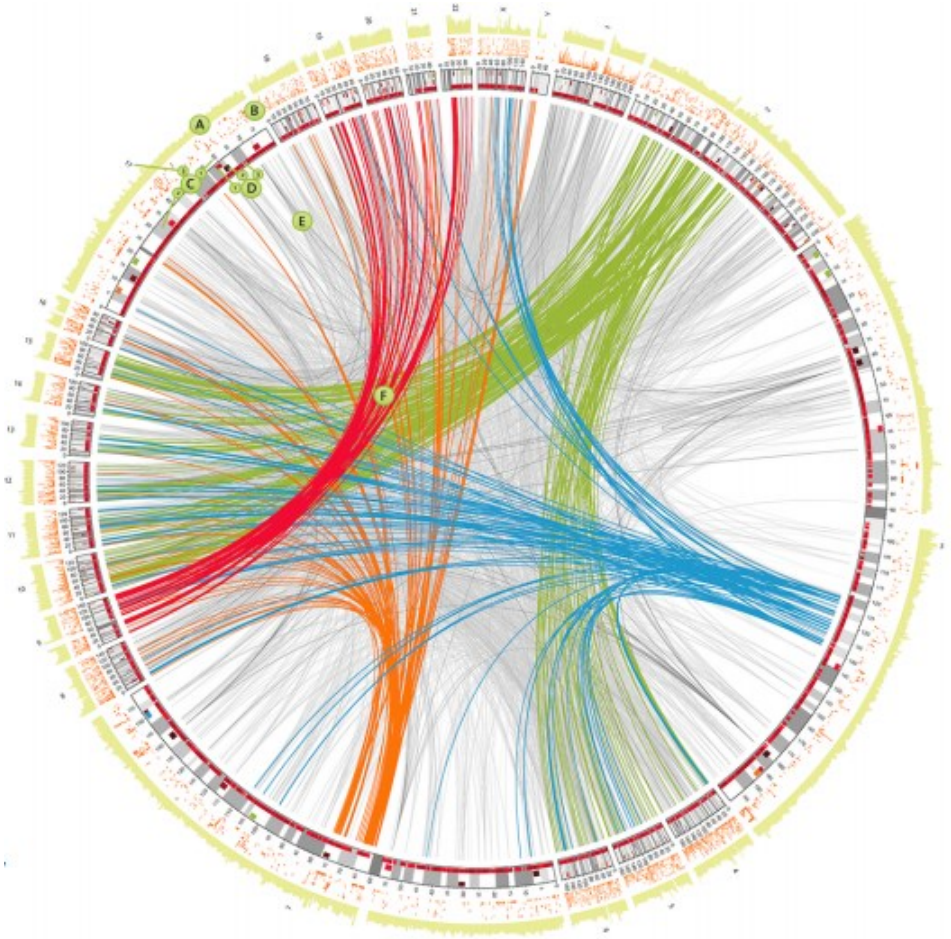
- Good for tree-like networks





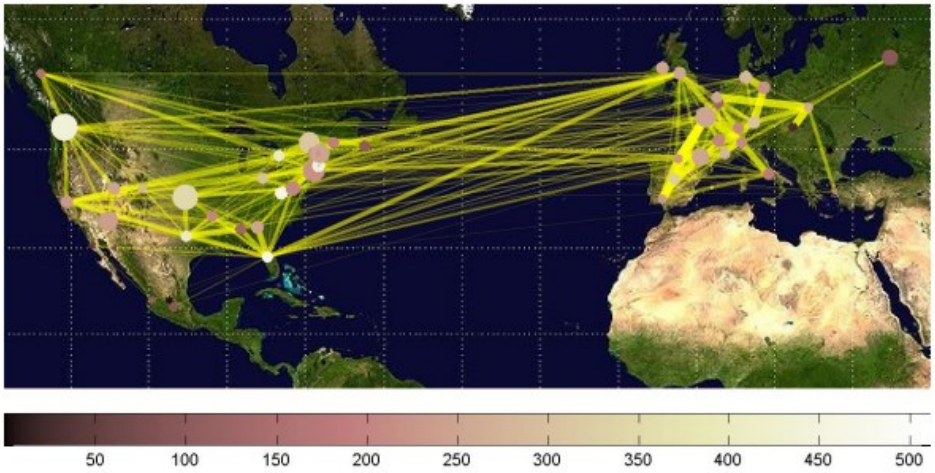
Circular Layout

- Nodes are positioned in a circle.
- Nodes with similar attributes may be grouped to see interactions between groups

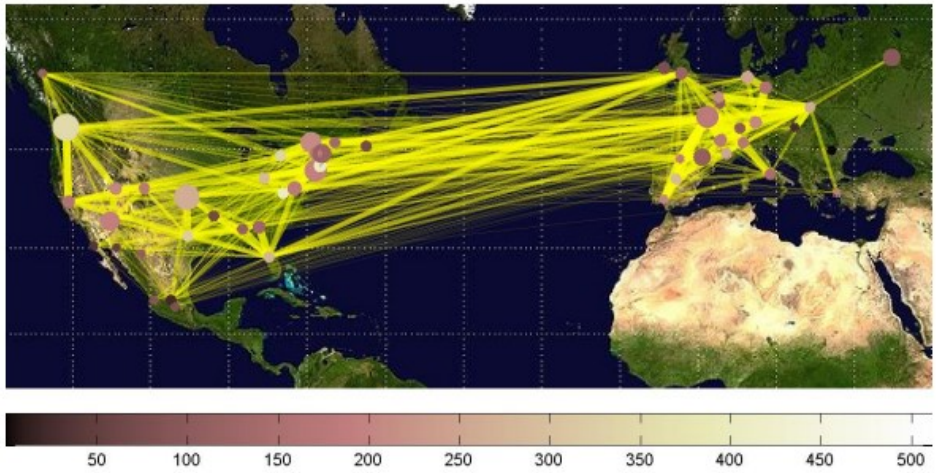


Network on Map

Figure 2a: Geographic map of buyer-supplier sub-network (the size of the bubble reflects location quotient, color gradient reflects the density of local linkages (normalized by the number of companies in the cluster) and the thickness of the line illustrates the density of trans-local linkages with other industrial clusters).



Buyer-supplier 2002-2005



Buyer-supplier 2010-2014