Formal Methods II
Graphs and Networks (Part I)

Rolf Pfeifer and Ruedi Fuechslin
Assistants:
Tobias Klauser, TA
Costas Dermitzakis, Qian Zhao
Fall term 2013
Motivation: networks

- Q: examples of networks?

Note: networks do not exist, but a part of reality can be - productively - viewed as a network.

Social networks
Information networks
Technological networks
Biological networks
(Newman, 2010)
Types of networks (1)

• Technological networks
  - internet
  - telephone networks
  - power grids
  - transportation networks
  - delivery and distribution networks (e.g. gas pipelines)

There is no generally accepted classification of networks, classifications are always arbitrary and they depend on the goals. Here is one from Newman’s book:

social networks
information networks
technological networks
biological networks

(Newman, 2010)
Types of networks (2)

- Social networks
  - friendship
  - business partners
  - sexual relations
  - scientific communities
  - hobbies
  - ...
Types of networks (3)

• Networks of information
  - WWW, p. 63 (Newman)
  - citation networks
  - other information networks (p2p, for sharing of files, recommender networks, keyword indices, relations between word classes in a thesaurus, etc.)
Types of networks (4)

- Biological networks
  - biochemical networks
    - metabolic networks
    - protein-protein interaction networks
    - genetic regulatory networks
- neural networks
- ecological networks (eat, parasitize, compete for resources, pollination, seed dispersal, etc.)
  - food webs
Interesting questions

- “distance in social networks” (“small worlds”)
- stability of computer networks (to terrorist attacks or disasters)
- scalability with size increase (times 1000)
- spread of mobile phone viruses (Barabasi et al., 2009)
- spread of viruses (H1N1 - the “bird flu”!!)
- breakdown of airline networks?
- power grids (e.g. SBB a few years ago)
- dynamics of genetic regulatory networks
References: scientific

- Newman, M.E.J. (2010). *Networks - An Introduction. Oxford University Press* (a general introduction to all aspects of networks; comprehensive; rather mathematical; at the moment, the only real textbook on network theory)

- Sporns, O. (2011). *Networks of the Brain. Cambridge, Mass.: MIT Press* (application of network theory to the understanding of the brain; short general introduction to network theory; recommended for anyone interested in neuroscience; written by a great neuroscientist)

- Sporns, O. ... The human connectome.


References: popular science

- Barabási, A.-L. (2002). Linked - How everything is connected to everything else and what it means for business, science, and everyday life. Penguin Books (entertaining, comprehensive, popular science introduction to networks with many informative examples, written by one of the top experts in the field)

Discovery of “small world” networks

Stanley Milgram’s experiment in 1960s: (famous for controversial experiment)

letters to random selection of people (Nebraska and Kansas)

please forward letter to stockbroker in Boston/ no address

send to someone you know and who might be socially closer to the stockbroker
Just for interest: Milgram’s authority experiment

- cover story: experiment on learning
- subjects can be seen; if they make a mistake, an electric shock is applied; on repetition of error, increase the voltage
- at some point: painful, some subjects start screaming
- experimenter encourages people to increase voltage, even when approaching lethal level

One of the astonishing insights was how easily subjects (people in general) submit to authorities, even if it is against their own convictions.
Discovery of “small world” networks

Stanley Milgram’s experiment in 1960s: (famous for controversial experiment)

letters to random selection of people (Nebraska and Kansas)

please forward letter to stockbroker in Boston/ no address

send to someone you know and who might be socially closer to the stockbroker

3/4 lost, 1/4 made it: in less than 6 steps! (42 of

seems to hold universally: Frankfurter Allgemeine — Kebap shop owner in Frankfurt and actor in Hollywood.

You and president Barrack Obama: Pfeifer — Josh Bongard — he knows president directly (Josh Bongard is the recipient of the PECAST, the Presidential Early Career Award for Science and Technology, which is handed over personally by the president).
How to take part in this study
(Milgram’s instructions)

1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.

2. DETACH ONE POSTCARD. FILL IT OUT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

3. IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.

4. IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POST-CARDS AND ALL) TO A PERSONAL
Kevin Bacon data base

- “Bacon number”: recall — Melanie Winiger: 2
distance 0: played in same movie (recall the
demonstration)

- Erdös number (famous mathematician):
distance 0: joint publication

Erdoes, a gifted mathematician with over 1500 published papers;
no home — staying over with math friends;
invention and study of random networks;
reason for random graphs/networks: nice mathematical properties;
no real interest in modeling real world;
rather: beauty of abstract things.
Random graphs are also used as a standard against which other graphs and networks are compared.
Königsberg is a town on the Preger River, which in the 18th century was German but now is Russian. Within the town there are two river islands that are connected to the banks with seven bridges.

It became a tradition to try to walk around the town in a way that only crossed each bridge once, but it proved to be a difficult problem. Leonhard Euler, a Swiss mathematician in the service of the Russian empress Catherine the Great, heard about the problem. In 1736 Euler proved that the walk was not possible to do. He proved this by inventing a kind of diagram called a network (or a graph) that is made up of vertices and arcs.
Königsberg is a town on the Preger River, which in the 18th century was a German town but now is Russian. Within the town there are two river islands that are connected to the banks with seven bridges. It became a tradition to try to walk around the town in a way that only crossed each bridge once, but it proved to be a difficult problem. Leonhard Euler, a Swiss mathematician in the service of the Russian empress Catherine the Great, heard about the problem. In 1736 Euler proved that the walk was not possible to do. He proved this by inventig a kind of diagram called a network (or a graph) that is made up of vertices and arcs. It can be clearly seen that all vertices are odd (have an odd number of arcs). To have an odd vertex, you would have to begin or end the trip at that vertex. Since there can only be one beginning and one end, there can only be two odd vertices if you’re going to be able to trace over each arc only once. Since the bridge problem has 4 odd vertices, it isn’t possible to do!
Graph theory

- study of graphs
- graph
- node/vertex (pl. vertices)
- edge/arc
- degree, in-degree, out-degree

graph: collection of vertices and edges
node/vertex: simple objects that can have names and other properties
edge/arc: connection between two vertices
(de standard definitions from graph theory)
degree, in-degree, out-degree: number of edges of vertex (or incoming; outgoing)
Graph theory

- distance
- path
- length of path
- adjacency matrix
- adjacency structure (with or without weight)
- distance matrix
- connected graph
- component
- bipartite graph
- cycle
- Hamilton path/cycle

**distance**: shortest path

**path**: ordered sequence of distinct nodes and links, linking a source node j to a target node i

**length of path**: number of distinct connections

**adjacency matrix**: for what kinds of graphs is this representation appropriate? —> dense graphs

**sparse graphs**: graphs with relatively few edges, e.g. less than V log V (V, number of vertices)
**dense graphs**: graphs with with relatively few of the possible edges missing

**adjacency structure**: for sparse graphs (done with linked lists)

**weighted networks**: simply put numbers into adjacency matrix - examples: strengths of social ties, synaptic strengths, bandwidth of data channels, number of lanes in a highway network, etc.

**distance matrix**: $d(i,j)$ distance between node j and node i

**connected graph**: a graph is called connected if given any two vertices i, j, there is a path from i to j.

**component**: a graph that is not connected can be divided into connected components (disjoint connected subgraphs).

**bipartite graph**: a graph is bipartite if its vertices can be partitioned into two disjoint subsets U and V such that each edge connects a vertex from U to one of V.

**cycle**: A cycle in a directed network is a closed loop of edges with the arrows on each of the edges pointing the same way around the loop.

In graph theory, an Eulerian path is a path in a graph which visits each edge exactly once. Necessary condition for Eulerian circuit: starts and ends in the same vertex

**cycle**: A cycle in a directed network is a closed loop of edges with the arrows on each of the edges pointing the same way around the loop.
**Network examples**

<table>
<thead>
<tr>
<th>network</th>
<th>vertex</th>
<th>edge</th>
</tr>
</thead>
<tbody>
<tr>
<td>internet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>www</td>
<td></td>
<td></td>
</tr>
<tr>
<td>citation networks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>power grid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>friendship network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>metabolic network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>neural network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>food web</td>
<td></td>
<td></td>
</tr>
<tr>
<td>genetic regulatory net</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Donnerstag, 28. November 13

see Newman, p. 110, Table 6.1
## Network examples

<table>
<thead>
<tr>
<th>network</th>
<th>vertex</th>
<th>edge</th>
</tr>
</thead>
<tbody>
<tr>
<td>internet</td>
<td>computer/router</td>
<td>cable or wireless</td>
</tr>
<tr>
<td>www</td>
<td>wep page</td>
<td>hyperlink</td>
</tr>
<tr>
<td>citation networks</td>
<td>article, patent</td>
<td>citation</td>
</tr>
<tr>
<td>power grid</td>
<td>generating station</td>
<td>transmission line</td>
</tr>
<tr>
<td>friendship network</td>
<td>person</td>
<td>friendship</td>
</tr>
<tr>
<td>metabolic network</td>
<td>metabolite</td>
<td>metabolic reaction</td>
</tr>
<tr>
<td>neural network</td>
<td>neuron</td>
<td>synapse</td>
</tr>
<tr>
<td>food web</td>
<td>species</td>
<td>predation</td>
</tr>
<tr>
<td>genetic regulatory net</td>
<td>genes</td>
<td>transcription factors</td>
</tr>
</tbody>
</table>
Graph theory: questions

- is there a path from a to b? reachability?
- shortest path?
- entire graph connected?
- converting graphs to trees (depth-first and breadth-first search)
Network theory: statistical properties of entire networks
Social networks
basic intuitions
Elementary network arithmetic: “warm-up”

- Aunt Mabel, 50 acquaintances
- each acquaintance, 50 acquaintances
- 1, 2, …, 5, 6 steps

2 steps: 2500
3 Steps …
5 Steps: 312'5000'000
6 Steps: 15'625'000'000 --> covers easily everybody on planet; explains 6 steps
Elementary network arithmetic: “warm-up”

- Aunt Mabel, 50 acquaintancies
- each acquaintance, 50 acquaintances
- 1, 2, ..., 5, 6 steps
- 5 Steps: 312'5000'000
- 6 Steps: 15'625'000'000 -> covers easily everybody on planet; explains 6 steps
- Q: problem with argument?

Donnerstag, 28. November 13

2 steps: 2500
3 Steps ...
5 Steps: 312'5000'000
6 Steps: 15'625'000'000 --> covers easily everybody on planet; explains 6 steps
Reply: not 50 DIFFERENT people
Elementary arithmetic: second example

- circle of $6 \times 10^9$ (people) (1D lattice, structured network)
- each person linked to 50 neighbors
- Q: degree of separation?
- Q: add 2 out of 10'000 random connections?
- dramatic collapse in number of steps

Degree of separation (2x50=10p2; half the circle)
$6 \times 10^9 / 10^2 = 6 \times 10^7 = 60 \times 10^6$ (60 million steps required, going 50 steps at a time around half the circle)
add 2: $60 \times 10^6 \rightarrow 8!!$
add 3: $\rightarrow 5!!$
local clustering remains the same
The power of “weak ties”

Network theory: basic concepts (1)

- average path length (characteristic path length)
- average degree of a node
- distribution of degrees
- power law distribution
- clustering coefficient/average clustering coefficient
- betweenness
- random network
- scale-free networks

Interest is in statistical/global properties of networks
Network theory: basic concepts

- average path length (characteristic path length)


Average path length, also called average degree of separation (or characteristic path length) is the mean over all the shortest path lengths in the network, i.e. for all pairs of nodes (i, j).

Sometimes the median is used instead of the means of the shortest path lengths (Watts, p. 29):
The characteristic path length (L) of a graph is the median of the means of the shortest path lengths connecting each vertex v in V(G) to all other vertices. That is, calculate d(v, j) for all j in V(G) and find dv(bar) for each v. Then define L as the median of {dv(bar)}.

reason for taking median rather than mean? --> get rid of distortions by a few extreme values, e.g. if there are only very few nodes with certain characteristics, they might strongly affect the mean, whereas the median is insensitive to extreme values.

Q: mean?
Q: median? (arrange data in ascending order, take the one in the middle - if even, take the mean of the two middle values)
Network theory: basic concepts

- average degree of a node

mean over all degrees in the network (indegree, outdegree)
Network theory: basic concepts

- distribution of degrees, cumulative degree distribution

often, we are not only interested in the average, but the distribution, i.e. how many nodes with low degrees, how many with high degrees, how many in between? (examples to follow in Watts and Strogatz model)

Cumulative degree distribution:
cumulative probability distribution of degrees, i.e., the fraction of vertices that have degree greater than or equal to k.
Network theory: basic concepts

- power law distribution

\[ f(x) = ax^k + \text{const} \]

scaling: \( f(cx) \) is proportional to \( f(x) \)
Scale invariance and power law

\[ f(x) = ax^k + o(x^k), \]

where \( a \) and \( k \) are constants, and \( o(x^k) \) is an asymptotically small function of \( x^k \).

Here, \( k \) is typically called the \textit{scaling exponent}, where the word "scaling" denotes the fact that a power-law function satisfies \( f(cx) \propto f(x) \) where \( c \) is a constant. Thus, a rescaling of the function’s argument changes the constant of proportionality but preserves the shape of the function itself. This point becomes clearer if we take the \textit{logarithm} of both sides:

\[ \log(f(x)) = k \log x + \log a. \]

Notice that this expression has the form of a \textit{linear relationship} with slope \( k \). Rescaling the argument produces a linear shift of the function up or down but leaves both the basic form and the slope \( k \) unchanged.

\[ f(x) = a.x^{**k} \]
\[ f(cx) = a.(cx)^{**k} = c^{**k} . a.x^{**k} \quad \text{prop. } f(x) \]
\[ \log(f(x)) = k \log(x) + \log(a) \quad \text{(straight line in } \log-\log \text{ paper)} \]
\[ \log(f(cx)) = \log (a.(cx)^{**k}) = \log(ac^{**k}) + k \log(x) \quad \text{(parallel to } \log(f(cx))) \]
Scale invariance and the power law

The main property of power laws that makes them interesting is their scale invariance. Given a relation \( f(x) = ax^k \), scaling the argument \( x \) by a constant factor causes only a proportionate scaling of the function itself. That is,

\[
f(cx) = a(cx)^k = c^k f(x) \propto f(x).
\]

That is, scaling by a constant simply multiplies the original power-law relation by the constant \( c^k \). Thus, it follows that all power laws with a particular scaling exponent are equivalent up to constant factors, since each is simply a scaled version of the others. This behavior is what produces the linear relationship when both logarithms are taken of both \( f(x) \) and \( x \), and the straight-line on the log-log plot is often called the signature of a power law. Notably, however, with real data,
Network theory: basic concepts

- clustering coefficient/average clustering coefficient

The clustering coefficient is defined as follows: Suppose that a vertex \( v \) has \( k_v \) neighbors. then at most \( k_{\text{max}} = \frac{k_v(k_v-1)}{2} \) edges can exist between them (this occurs when every neighbor of \( v \) is connected to every other neighbor of \( v \)). Let \( C_v \) denote the fraction of these allowable edges that actually exist. Define \( C \) as the average of \( C_v \) over all \( v \).

Intuitively, in a friendship network, this coefficient measures to what extent my friends are also friends of each other.

\[ C_v = \frac{k_{\text{actual}}}{k_{\text{max}}} \]
Network theory: basic concepts

- clustering coefficient/average clustering coefficient - Example (from Watts and Strogatz)

```
Clustering coefficient:
in this network (with a lattice structure): for k=2 --> 4 neighbors, kmax=6, kactual=3, C=0.5
```
Network theory: basic concepts

- betweenness/centrality

The number of shortest paths going through a particular node. This number is used, for example, to characterize airports in a flight network (see case study, below).
Network theory: basic concepts

• random network

A network with uniform connection probabilities (and a binomial degree distribution). All nodes have roughly the same degree (single scale).

In probability theory and statistics, the binomial distribution is the discrete probability distribution of the number of successes in a sequence of $n$ independent yes/no experiments, each of which yields success with probability $p$. Such a success/failure experiment is also called a Bernoulli experiment or Bernoulli trial; when $n = 1$, the binomial distribution is a Bernoulli distribution. The binomial distribution is the basis for the popular binomial test of statistical significance.

(for details, see Newman 2010, p. 401-402).

In practice: typically networks are not random - they form an interesting object of mathematical studies and they are used in comparisons to other kinds of networks.

Scale-free networks (see next slide) are much more relevant for real-world networks.
Network theory: basic concepts

- scale-free network

Power law distribution is scale free (as shown above)

“Scale free” means that degrees are not grouped around one characteristic average degree, but can spread over a wide range of values, often spanning several orders of magnitude.

Scale-free networks are ubiquitous in the real world (earthquakes, avalanches, word frequencies, WWW, Internet, etc.)
Network theory: basic concepts

• aristocratic network

because of the highly uneven degree distribution, they are sometimes called “aristocratic” examples: Internet, WWW: Many nodes with low degree, few nodes with high degree.
Network theory: basic concepts

- egalitarian network

In these types of networks, the nodes have roughly equal degree. Examples: grids.
Formal network analysis
The Watts and Strogatz model
Collective dynamics of “small-world” networks

Figure 1 Random rewiring procedure for interpolating between a regular ring lattice and a random network, without altering the number of vertices or edges in the graph. We start with a ring of $n$ vertices, each connected to its $k$ nearest neighbors by undirected edges. (For clarity, $n = 20$ and $k = 4$ in the schematic examples shown here, but much larger $n$ and $k$ are used in the rest of this publication.) We choose a vertex and the edge that connects it to its nearest neighbor in a clockwise sense. With probability $p$, we reconnect this edge to a vertex chosen uniformly at random over the entire ring, with duplicate edges forbidden; otherwise we leave the edge in place. We repeat this process by moving clockwise around the ring, considering each vertex in turn until one lap is completed. Next, we consider the edges that connect vertices to their second-nearest neighbors clockwise. As before, we randomly rewire each of these edges with probability $p$, and continue this process, circulating around the ring and proceeding outward to more distant neighbors after each lap, until each edge in the original lattice has been considered once. (As there are $nk/2$ edges in the entire graph, the rewiring process stops after $k/2$ laps.) Three realizations of this process are shown, for different values of $p$. For $p = 0$, the original ring is unchanged; as $p$ increases, the graph becomes increasingly disordered until for $p = 1$, all edges are rewired randomly. One of our main results is that for intermediate values of $p$, the graph is a small-world network: highly clustered like a regular graph, yet with small characteristic path length, like a random graph.
Characteristic path length and clustering coefficient

Characteristic path length $L(p)$ and clustering coefficient $C(p)$ for the family of randomly rewired graphs described in Fig. 1. Here $L$ is defined as the number of edges in the shortest path between two vertices, averaged over all pairs of vertices. The clustering coefficient $C(p)$ is defined as follows. Suppose that a vertex $v$ has $k_v$ neighbors; then at most $k_v(k_v - 1)/2$ edges can exist between them (this occurs when every neighbor of $v$ is connected to every other neighbor of $v$). Let $C_v$ denote the fraction of these allowable edges that actually exist. Define $C$ as the average of $C_v$ over all $v$. For friendship networks, these statistics have intuitive meanings: $L$ is the average number of friendships in the shortest chain connecting two people; $C_v$ reflects the extent to which friends of $v$ are also friends of each other; and thus $C$ measures the cliquishness of a typical friendship circle. The data shown in the figure are averages over 20 random realizations of the rewiring process described in Fig. 1, and have been normalized by the values $L(0)$, $C(0)$ for a regular lattice. All the graphs have $n = 1000$ vertices and an average degree of $k = 10$ edges per vertex. We note that a logarithmic horizontal scale has been used to resolve the rapid drop in $L(p)$, corresponding to the onset of the small-world phenomenon. During this drop, $C(p)$ remains almost constant at its value for the regular lattice, indicating that the transition to a small world is almost undetectable at the local level.
Re-wiring demo


1. Erdos Renyi (random network)
2. Barabasi Albart (scale free)
3. Watts Strogatz (small world)

The applet is quite simple to show,

(about 2. scale free network)
If you want to show the power law of the scale-free network, click "setup BA" and then click "go". You can see nodes added to the right window and also the power low figure on the most right figure.

(about 3. small world)
As for small world network demonstration, please click "show WS" and click "rewire all". This time the applet describes the clustering coefficient and the average path length.
Examples of degree distributions

- (see next slide)
Cumulative degree distributions for six different networks. The horizontal axis for each panel is vertex degree $k$ (or in-degree for the citation and Web networks, which are directed) and the vertical axis is the cumulative probability distribution of degrees, i.e., the fraction of vertices that have degree greater than or equal to $k$. The networks shown are: (a) the collaboration network of mathematicians [182]; (b) citations between 1981 and 1997 to all papers cataloged by the Institute for Scientific Information [351]; (c) a 300 million vertex subset of the World Wide Web, circa 1999 [74]; (d) the Internet at the level of autonomous systems, April 1999 [86]; (e) the power grid of the western United States [416]; (f) the interaction network of proteins in the metabolism of the yeast S. Cerevisiae [212]. Of these networks, three of them, (c), (d) and (f), appear to have power-law degree distributions, as indicated by their approximately straight-line forms on the doubly logarithmic scales, and one (b) has a power-law tail but deviates markedly from power-law behavior for small degree. Network (e) has an exponential degree distribution (note the log-linear scales used in this panel) and network (a) appears to have a truncated power-law degree distribution of some type, or possibly two separate power-law regimes with different exponents.
An attempt to characterize networks in terms of three dimensions, randomness, modularity and heterogeneity (of node degree).
SF: scale-free
ER: Erdoes Renyi random graphs
Macaque cerebral cortex. Adjacency matrix and degree distribution.

An example of an adjacency matrix and a degree distribution. (A) The adjacency matrix records the presence (black square) and absence (white square) of corticocortical connections between regions of the macaque cortex. Many of the connections are symmetrical, and two main modules, corresponding to mostly visual (M1) and mostly somatomotor regions (M2), are indicated in the anatomical surface plot at the upper right. (B) The degree distribution (indegree plus outdegree for each node) is broad, with degrees ranging from 3 to 42.
This metabolic network has an unbelievable level of complexity.
Case study: Airline transport network — “betweenness”

Large-scale structure

- $S = 3'883$ cities
- 27,051 distinct city pairs
- $d(\text{Asia/Middle East}) = 3.5$ (see slide on community structure)
- $d(\text{worldwide}) = 4.4$
- Q: what does this mean?
- $d$ grows with log $S$
- Farthest cities in network: Mount Pleasant (Falkland Islands) to Wasu, Papua New Guinea: 15 different flights
Network theory: basic concepts

- betweenness: the number of shortest paths going through a particular node. This number is used, for example, to characterize airports in a flight network. Anchorage has relatively low degree but high betweenness.
Degree and betweenness distributions of the worldwide air transportation network.

(note: the cumulative degree distribution \( P(>k) \) gives the probability that a city has \( k \) or more connections to other cities;

\[ P(>k) = \text{sum over } k' = k \text{ to infinity of } p(k'), \]

where \( p(k') \) is the probability density function that the node has \( k' \) connections to other nodes).

(a) Cumulative degree distribution plotted in double-logarithmic scale. The degree \( k \) is scaled by the average degree \( z \) of the network. The distribution displays a truncated power-law behavior with exponent \( \alpha = 1.0 \pm 0.1 \)

(b) Cumulative distribution of normalized betweennesses plotted in double-logarithmic scale. The distribution displays a truncated power-law behavior with exponent \( \nu = 0.9 \pm 0.1 \). For a randomized network with exactly the same degree distribution as the original air transportation network, the betweenness distribution decays with an exponent \( \nu = 1.5 \pm 0.1 \). A comparison of the two cases clearly shows the existence of an excessive number of large betweenness values in the air transportation network.
Most-connected vs. most-central cities

25 most connected cities in the world

25 most central cities in the world

(a) Betweenness as a function of degree. For random networks (dashed line): quadratic function (gray region).

By contrast: airline transportation network:
- Many cities with high degree (connectedness), but low betweenness (centrality) (blue region: 25 most central cities).
- Many cities have small degree and large betweenness (centrality) (yellow region: 25 most connected cities).
Community structure

each dot represents a location, each color a community

modularity
high betweenness: connection between communities
end of Part I

Stay tuned for Part II!