Economic Networks

Theory and Applications

Giorgio Fagiolo
Laboratory of Economics and Management (LEM)
Sant'Anna School of Advanced Studies, Pisa, Italy

http://www.lem.sssup.it/fagiolo/
giorgio.fagiolo@sssup.it

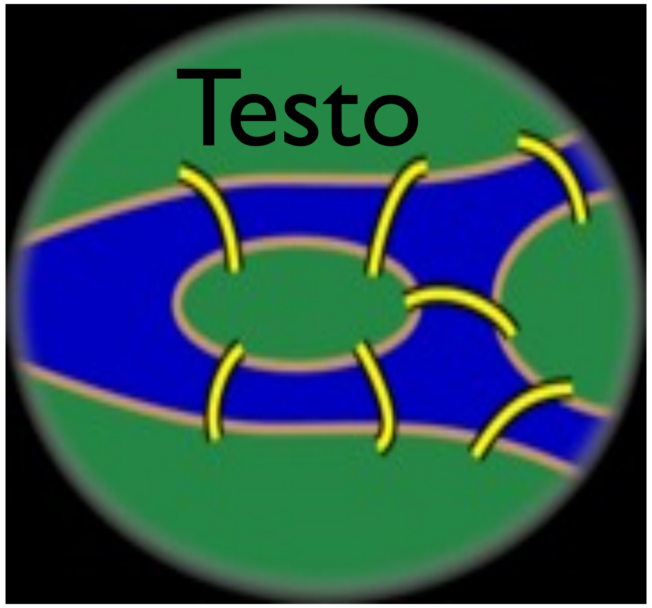
Networks and Graphs

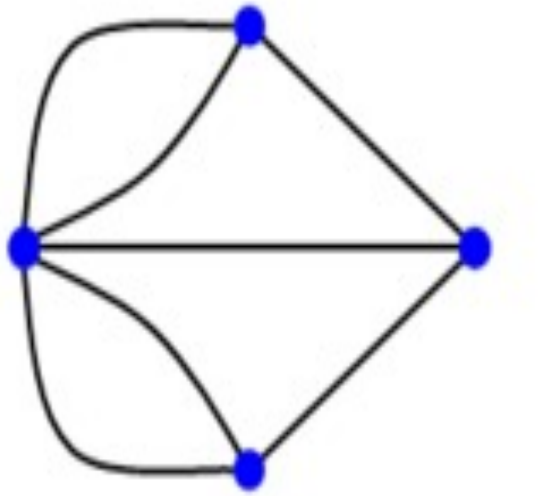
What's Next

- What is a network? Examples of networks
- Why networks are important for economists?
- Networks and graphs
- Measures and metrics on networks
- Distributions of metrics and measures in large networks
- Modeling Networks
- Economic applications

Graph Theory: History

Leonhard Euler's paper on “Seven Bridges of Königsberg”, published in 1736.





Problem of the Königsberg (now Kaliningrad, Russia) bridges: find a walk through the city that would cross each bridge once and only once

Graph Theory: Definitions

- A **graph** G is a pair G=(V, E)
- V is the set of **nodes** or **vertices** {1,...,N}
- E is a set of **edges** or **links**, where elements of V are (unordered) pairs {u,v}, where $u,v \in V$
- **Order** of the graph = |V|=N
- **Size** of the graph = |E| = L (but often size of a graph = N)
- **Subgraph**: a graph $H=(V_H, E_H)$ of another graph $G=(V_G, E_G)$ if $V_H \subseteq V_G$ and $E_H \subseteq E_G$
- **Induced subgraph** of G: is a subgraph G'=(V',E') where $V' \subseteq V$ is a pre-specified subset of vertices and $E' \subseteq E$ is the collection of edges in G among vertices in V'

Graph Theory: Multigraphs and Digraphs

- **Multigraphs**, i.e graphs containing:
  - ✓ Replications of the same edge in E (multi-edges)
  - ✓ Possibly: Edges starting and ending in the same vertex (loops)
- **Simple graphs**: graphs without loops and multi-edges
- **Undirected graphs**: graphs where each edge in E has no ordering on its vertices
- **Directed graphs** (binary digraphs or simply digraphs): graphs where E has an ordering: {u,v}≠{v,u}. Edges in a digraph are called "arcs"
- The **underlying graph** of a digraph G is the graph obtained by removing directionality
- Given a digraph, it is possible to have up to two arcs between any pair of nodes (u,v) without the digraph being a multi-digraph. When given (u,v), {u,v}∈E and {v,u}∈E, then the two arcs are said to be "mutual" or "reciprocated"

Graph Theory: Connectivity (I)

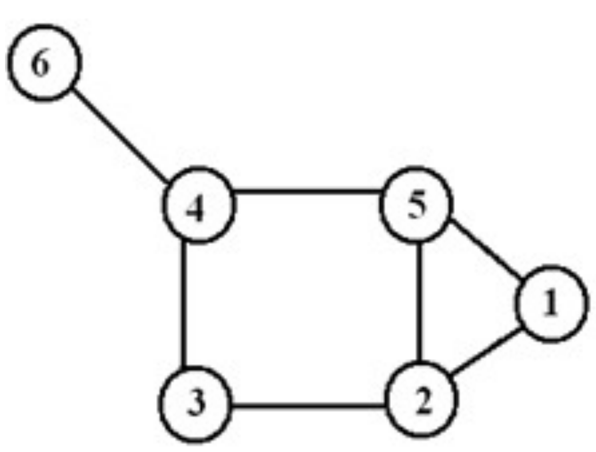
- **Adjacency**:
  - ✓ two nodes (u,v)∈V are adjacent in G=(V,E) if they are joined by an edge
  - ✓ two edges (e_1,e_2) are adjacent in G=(V,E) if they have a common endpoint
- A node v∈V is **incident** on an edge e∈E if v is an endpoint of e
- **Node degree** (ND): the degree d(v) of a node v is the number of edges incident on v
- The **degree sequence** of a graph G is the sequence formed by arranging all NDs in non-decreasing order
- Result: the sum of all NDs is equal to 2*L (prove it)

Graph Theory: Connectivity (II)

- ND in digraphs
  - ✓ **In-degree** of a node $d_{in}(v)$ = number of edges pointing in towards the node
  - ✓ **Out-degree** of a node $d_{out}(v)$ = number of edges pointing out from the node
- Digraphs have both a in-degree and out-degree sequence
- Result: sum of all $d_{in}(v)$ is equal to sum of all $d_{out}(v)$ (prove it)

Graph Theory: Moving on a Graph (I)

- **Walk** on G: a walk from v_0 to v_k in G is a sequence $\{v_0,e_1,v_1,\ldots,v_{k-1},e_k,v_k\}$ where the endpoints of e_i are $\{v_{i-1},v_i\}$, i.e. any possible route from v_0 to v_k (one can pass through the same edge many times)
- Walks can also be described by the ordered sequence of nodes: $\{v_0,e_1,v_1,\ldots,v_{k-1},e_k,v_k\}= \{v_0,v_1,\ldots,v_{k-1},v_k\}$
- The **length** of a walk is equal to the number of edges in the walk
- A walk is closed if $v_0 \equiv v_k$
- **Path**: walks without repeated vertices (apart from only initial and ending notes)

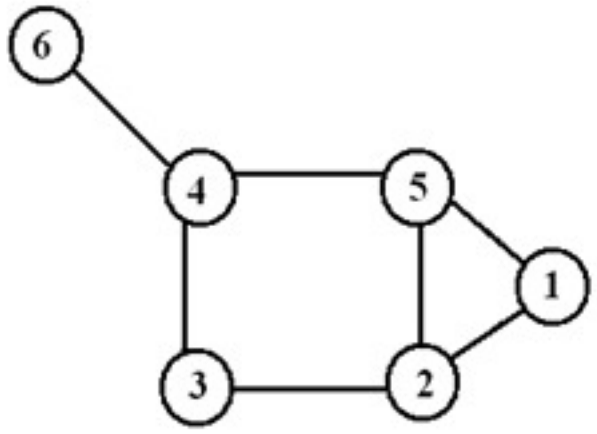


Walks and Paths

(1,2,5,2,3,4) walk of length 5
(1,2,5,2,3,2,1) closed walk of length 6
(1,2,3,4,6) path of length 4

Graph Theory: Moving on a Graph (II)

- **Cycle**: a path of length at least 3 for which the beginning and ending vertices are the same
- Graphs containing no cycles are called **acyclic**
- Walks and paths in digraphs extend naturally. E.g. a directed walk from v_0 to v_k proceeds from tail to head of arcs between v_0 to v_k
- A vertex v is **reachable** from another vertex u if there exists a path from v to u



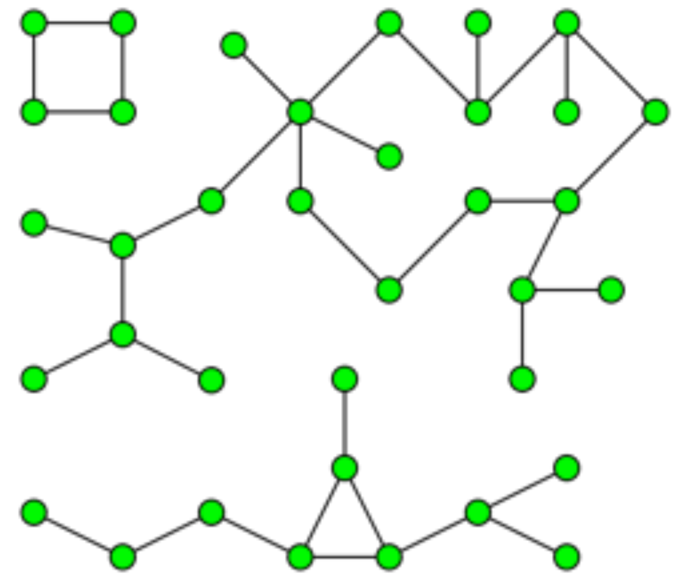
1,2,5,1 3-cycle

2,3,4,5,2 4-cycle

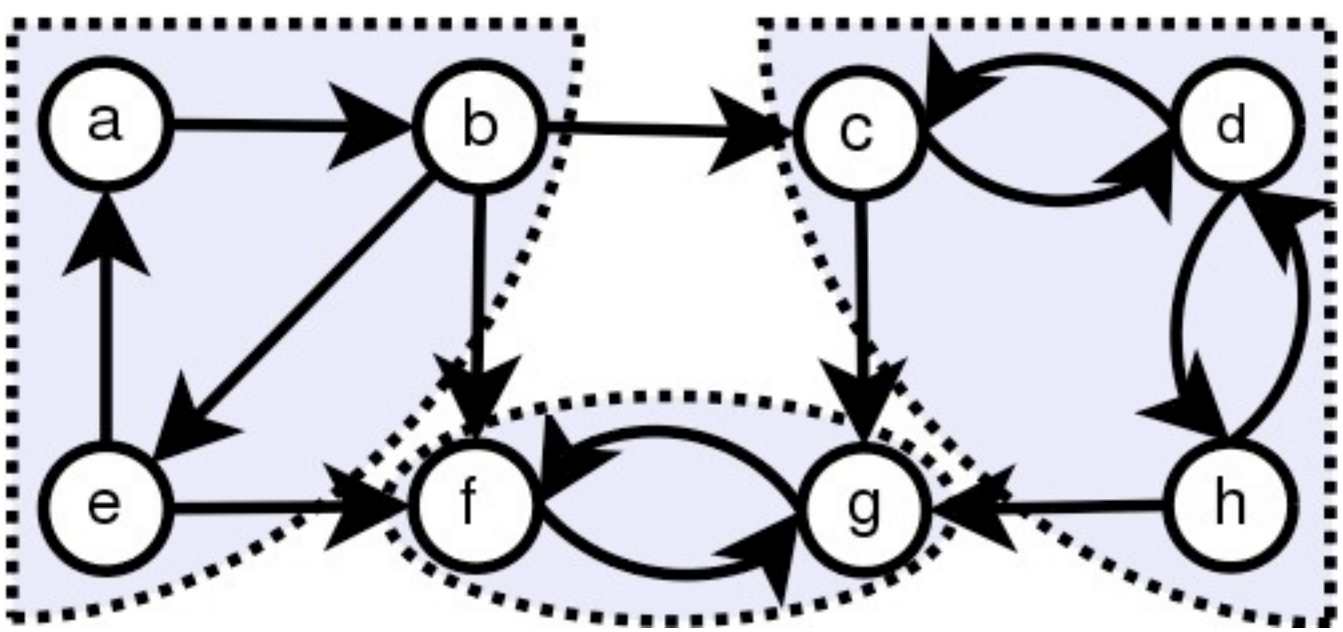
- A graph is said to be **connected** if every vertex is reachable from every other (i.e. if there exists a path between all pairs of vertices)
- A (connected) **component** of a graph is a maximally connected subgraph, i.e. a connected subgraph of G for which the addition of any other remaining node in G makes it not connected

Graph Theory: Moving on a Graph (II)

- Example: a graph with three connected components

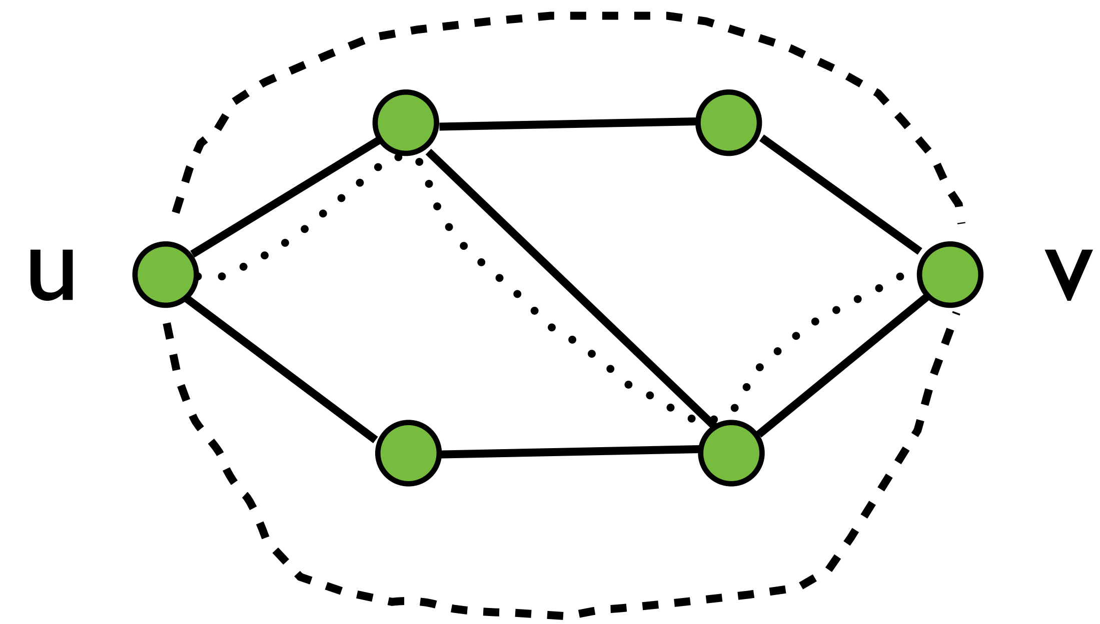


- A graph that is itself connected has exactly one connected component, consisting of the whole graph
- Connectivity does not naturally extend to digraphs. We need two additional concepts.
- A digraph G is **weakly connected** if its underlying graph is connected
- A digraph G is **strongly connected** if every node v is reachable from every u by a directed path



Graph Theory: Distances

- **Distance** between nodes: the distance between any two nodes (u,v) in a graph G is the length of the shortest path between (u,v), also called **geodesic distance**.
- The distance between any two nodes that are not connected by any path is typically set to be infinite
- The value of the longest distance in a graph (i.e. the maximum of all geodesic distances) is called **diameter** of a graph
- Geodesic paths need not to be unique. Example: (u,v) connected by 3 geodesics

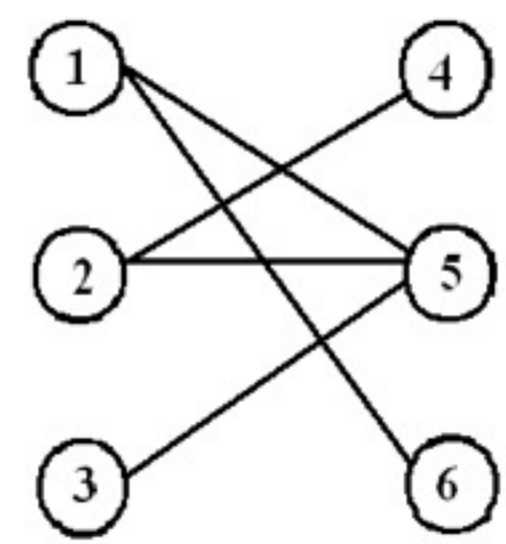


Graph Theory: Beyond Basic Graphs

- A **weighted** graph is a graph where to each edge $e=\{u,v\}$ is attached a (typically positive) weight w_{uv}. Unweighted graphs (or the binary counterpart of a weighted graph) can be defined setting $e \in E$ iff $w_{uv} \neq 0$. A binary graph can be considered as a particular case of a weighted graph, such that $w_{uv}=1$ iff $e \in E$ and $w_{uv}=0$ iff $e \notin E$

- A graph is a **k-mode** graph if nodes belong to k different sets. A **k-partite** graph is a k-mode graph where a link can only exist between two nodes if they belong to different subsets.

- Example. **Bipartite graph**: is a graph $G=(V,E)$ such that $V=\{V_1,V_2\}$ with $V_1 \cap V_2=\varnothing$ and $V_1 \cup V_2=V$, and each edge in E has one endpoint in V_1 and the other in V_2



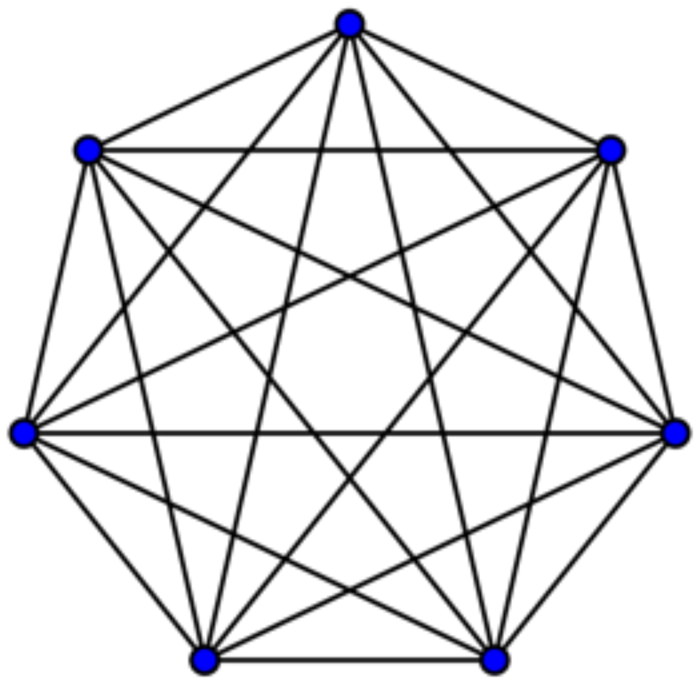
- **Weighted digraphs**: Digraphs where arcs are weighted. Notice that two arcs insisting on the same nodes can have different weights, i.e. $w_{uv} \neq w_{vu}$

- **Weighted multi-digraphs**: Weighted multi-digraphs where between any two nodes there may be many weighted arcs (arcs with different colors)

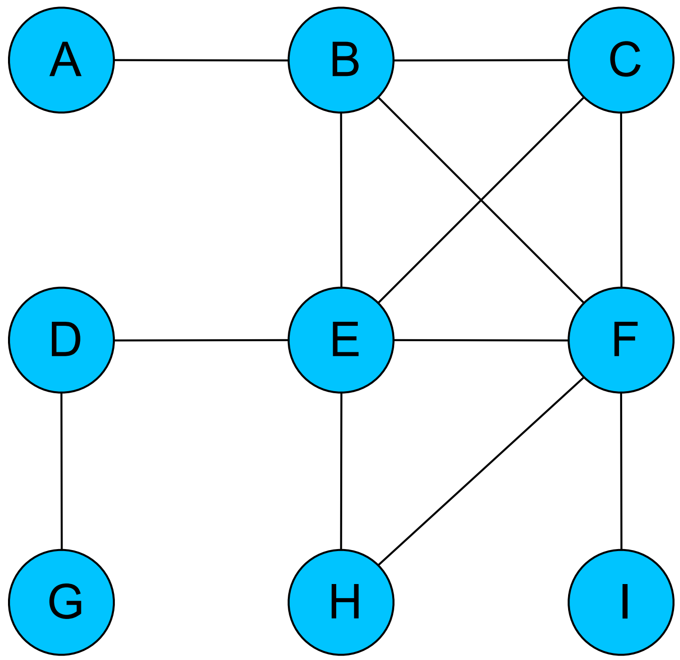
Graph Theory: Families of Graphs (1)

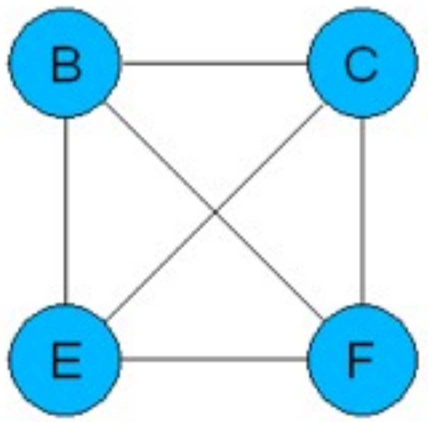
- **Complete graphs**: graphs where every node is joined to every other node. A simple graph with order N is complete if it has L=N(N-1)/2 edges [L=N(N-1) for simple digraphs]. Check it.

- A **clique** is a complete subgraph. A subgraph H of G is a maximal clique if it is complete and no other such subgraph contains it.



A complete (regular) graph

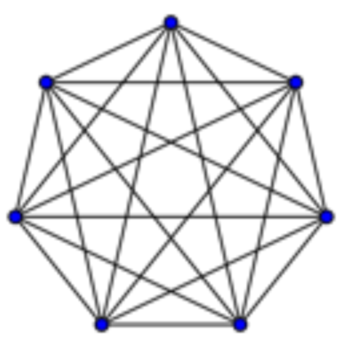




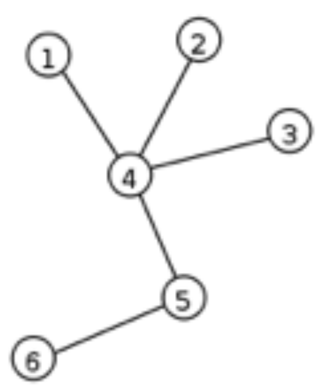
A clique (right) of a graph (left)

Graph Theory: Families of Graphs (2)

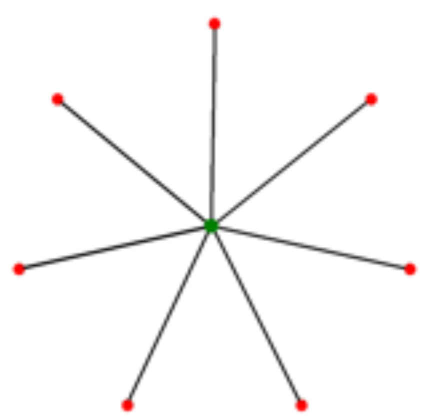
- A **d-regular graph** is a graph where every node has the same degree d. A particular type of d-regular graph is a **lattice**.
- A connected graph with no cycles is called a **tree**. A digraph whose underlying graph is a tree is called a directed tree. A **star** is a tree composed of a node attached to the N-1 other nodes by a single edge.
- A directed acyclic graph (**DAG**) is a digraph without cycles. The underlying graph of a DAG needs not to be a tree. Prove it with a counter example.



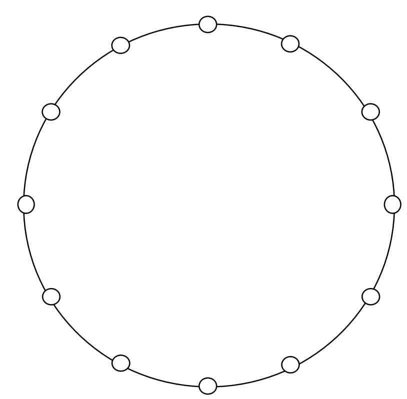
A complete (regular) graph



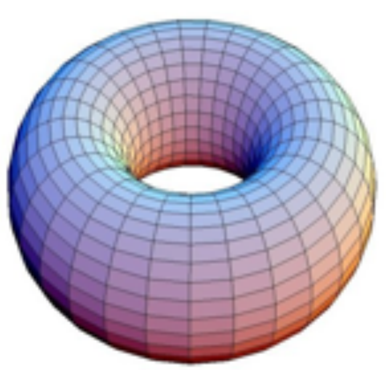
A tree



A star

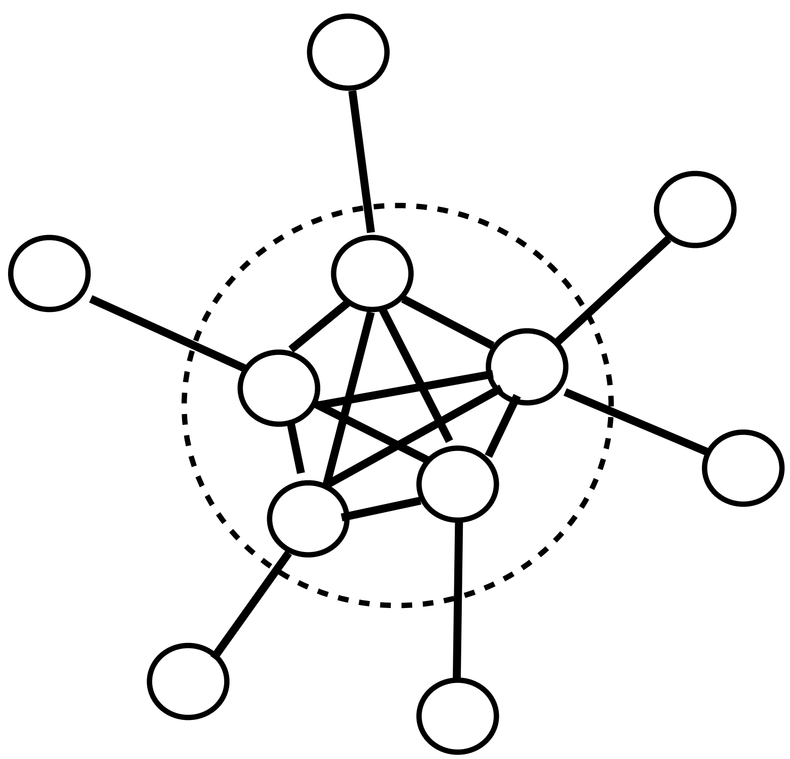


A 1-dim (regular) lattice

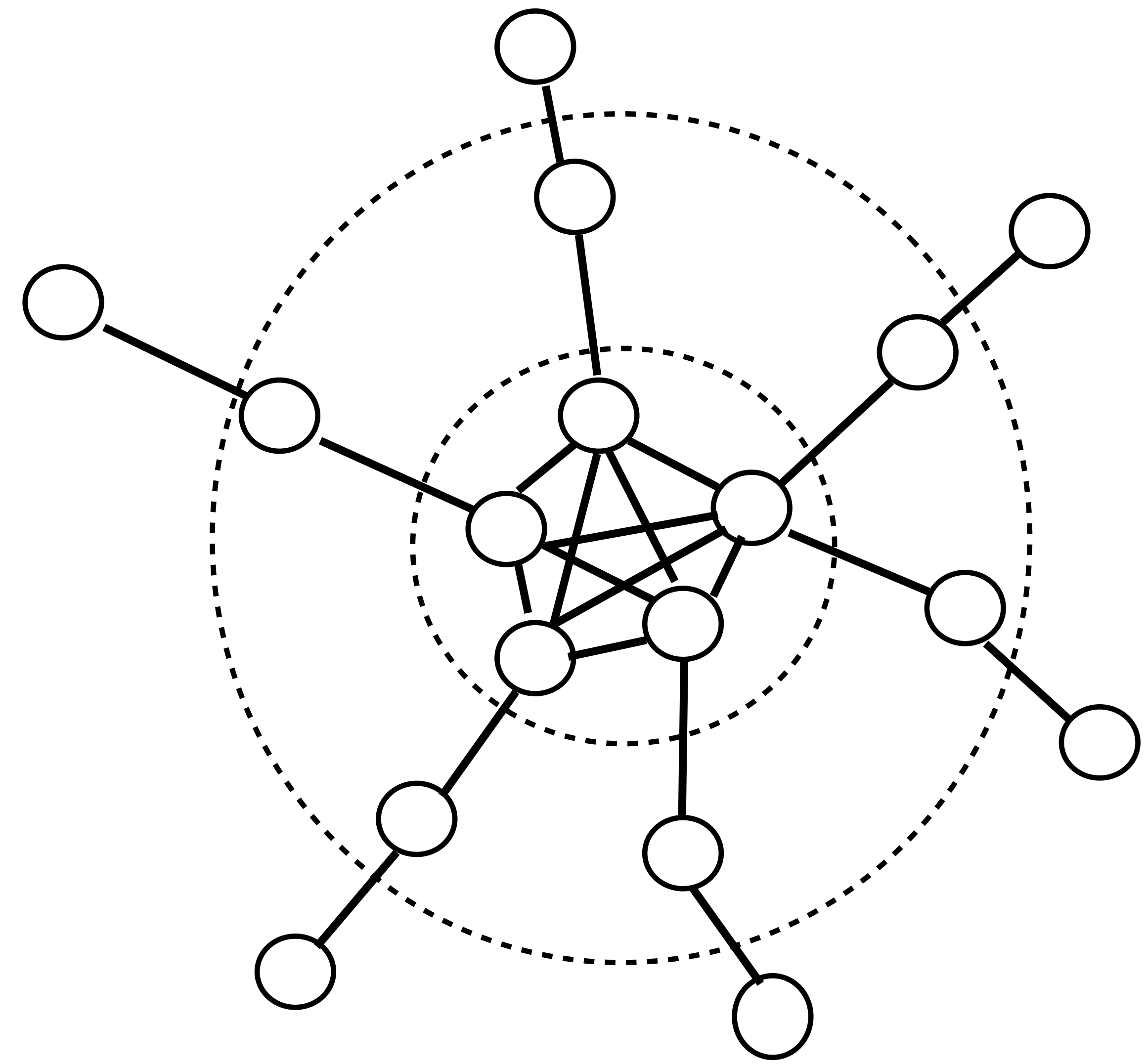


A 2-dim (regular) lattice

Core-Periphery and Tiered Networks



Core-Peryphery:
Dense core with
periphery layers



Tiered: Core,
intermediate
layer, periphery
layer

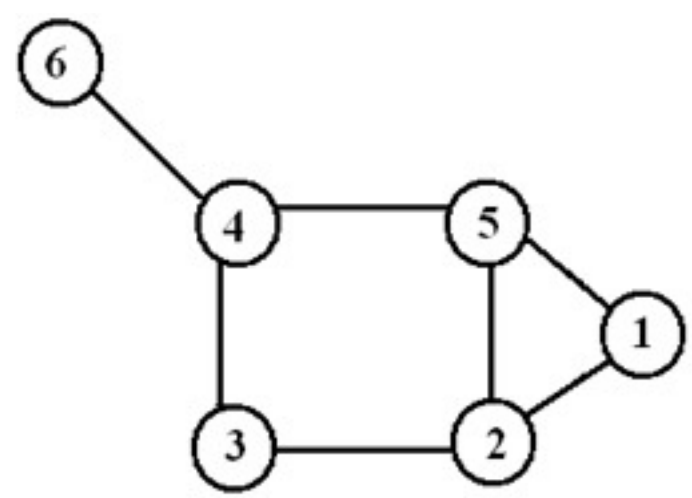
Mathematical Representations of a Graph (1)

- A simple undirected, binary graph G=(V,E) may be represented mathematically in two ways.

  1. **Matrix** representation (useful when N is small)
  2. **Edge-List** representation (useful when N is large and L<<N(N-1)/2, i.e. the graph is sparse)

	1	2	3	4	5	6
1	0	1	0	0	1	0
2	1	0	1	0	1	0
3	0	1	0	1	0	0
4	0	0	1	0	1	1
5	1	1	0	1	0	0
6	0	0	0	1	0	0

Adjacency Matrix



G=(V,E)

(1,2) 1
(1,5) 1
(2,5) 1
(2,3) 1
(3,4) 1
(4,5) 1
(4,6) 1

Edge List

- **Adjacency matrix**: a NxN matrix A, N=|V|, where a(i,j)=1 iff {i,j}∈E, and zero otherwise. Note: a(i,i)=0,∀i

- **Edge list**: a list of L=|E| rows, each row lists links and the corresponding weights (=1 in the case of a unweighted graph). Note: the edge list is equivalent to a sparse representation of A

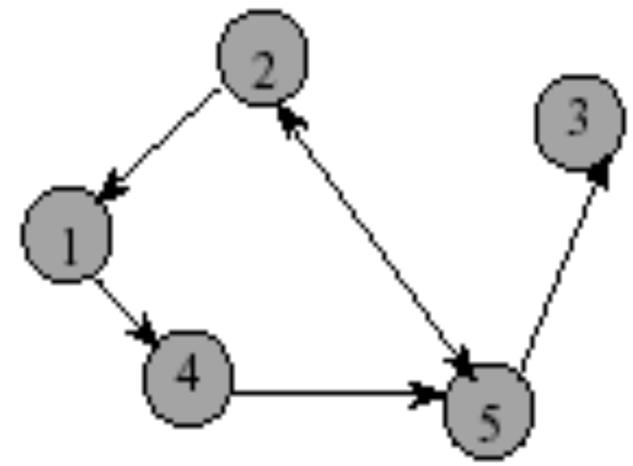
Mathematical Representations of a Graph (2)

- **Some useful results**:
  - ✓ If the graph is undirected, A is symmetric (check it)
  - ✓ The sum of row i's (or, equivalently, if G is undirected, column i's) entries of A gives the degree of node i (see before)
  - ✓ If we let $Z=A^r$ the r-th power of A, then the entry z(i,j) gives the number of paths of length r between i and j (prove it in the case of r=2). What does z(i,i) measure?

- **Extensions**:
  - ✓ If G is a digraph, then A is asymmetric. The list representation simply enumerates all links in place between ordered pairs (i.e., it may be the case that in the list both (i,j) and (j,i) appear)
  - ✓ If G is a weighted graph, the adjacency matrix becomes a **weight matrix W**: entries w(i,j) represent the weight of a link in place between i and j (and zero otherwise). If G is undirected, then W is symmetric.
  - ✓ If G is a weighted digraph, then W is asymmetric. The edge list contains all directed edges together with the corresponding weight in the third column (replacing the ones)

Mathematical Representations of a Graph (3)

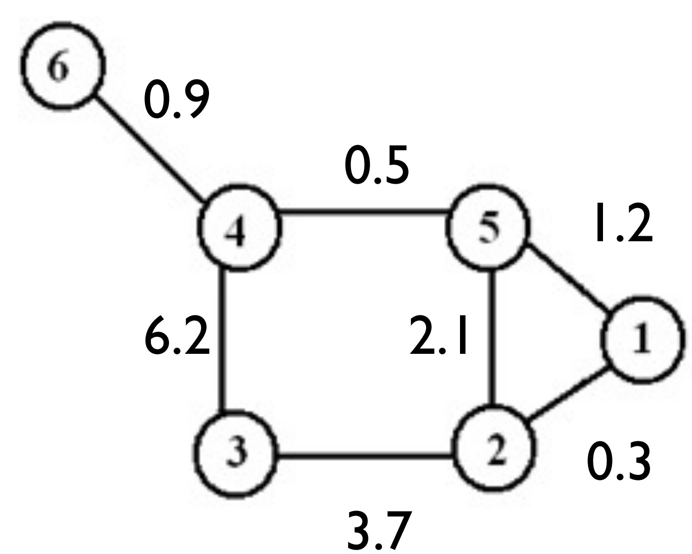
	1	2	3	4	5
1	0	0	0	1	0
2	1	0	0	0	1
3	0	0	0	0	0
4	0	0	0	0	1
5	0	1	1	0	0



(1,4) 1
(2,1) 1
(2,5) 1
(4,5) 1
(5,2) 1
(5,3) 1

- **A useful result**: If the graph G is a binary digraph, then
  - ✓ the sum of row i of A gives the node out-degree of i;
  - ✓ the sum of column j of A gives the node in-degree of j

W	1	2	3	4	5	6
1	0	0.3	0	0	1.2	0
2	0.3	0	3.7	0	2.1	0
3	0	3.7	0	6.2	0	0
4	0	0	6.2	0	0.5	0.9
5	1.2	2.1	0	0.5	0	0
6	0	0	0	0.9	0	0



(1,5) 1.2
(1,2) 0.3
(2,5) 2.1
(2,3) 3.7
(3,4) 6.2
(4,5) 0.5
(4,6) 0.9

Directionality and Symmetry

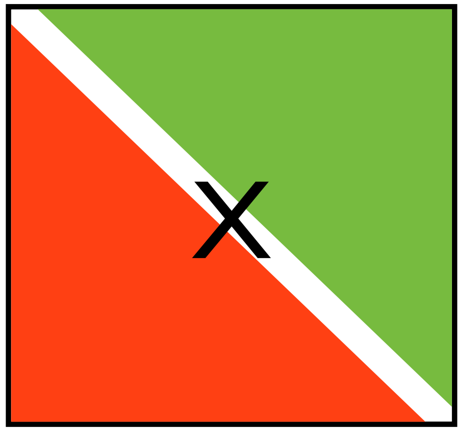
- A simple graph is undirected if its adjacency or weight matrix is symmetric, i.e. $A=A^T$ or $W=W^T$

- How can we measure "how much asymmetry" is present in a network?

  ✓ Computing the fraction of all present links that are reciprocated, i.e. the number of all cycles of length 2 divided by the number of all links: $tr(A^2)/L$ (prove it)

  ✓ Computing the correlation coefficient between the upper diagonal part and the lower diagonal part of the adjacency/weight matrix

  ✓ More sophisticated indicators based on $||A-A^T||^2$, see Fagiolo (2007)



$corr(x_{ij},x_{ij})$

- What if the network is sufficiently symmetric? We can symmetrize it.

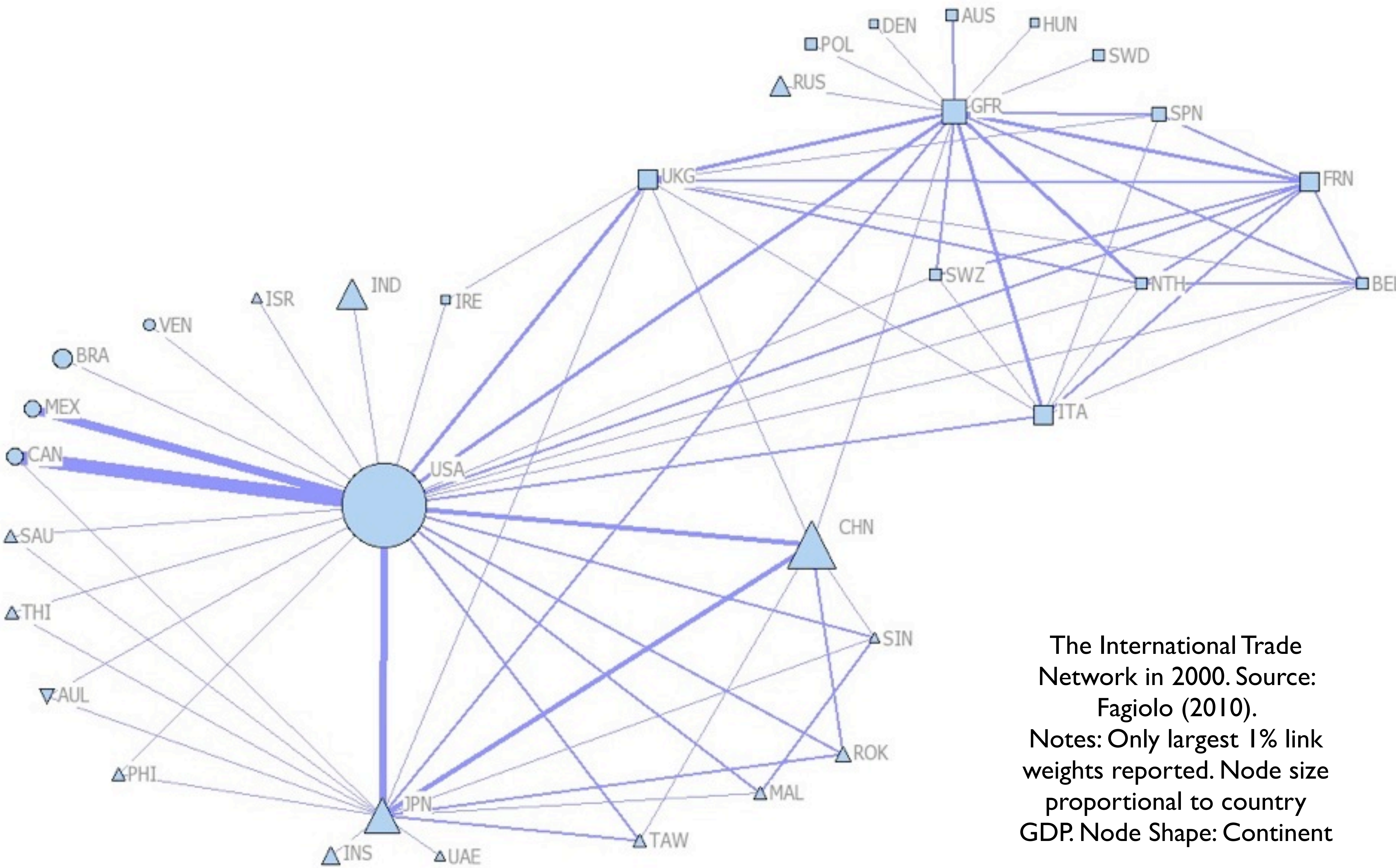
  ✓ Binary network: Replace a_{ij} with $\max\{a_{ij},a_{ji}\}$

  ✓ Weighted network: Replace w_{ij} with $½(w_{ij}+w_{ji})$ or $\sqrt{(w_{ij} \cdot w_{ji})}$

Visualization of Network Data

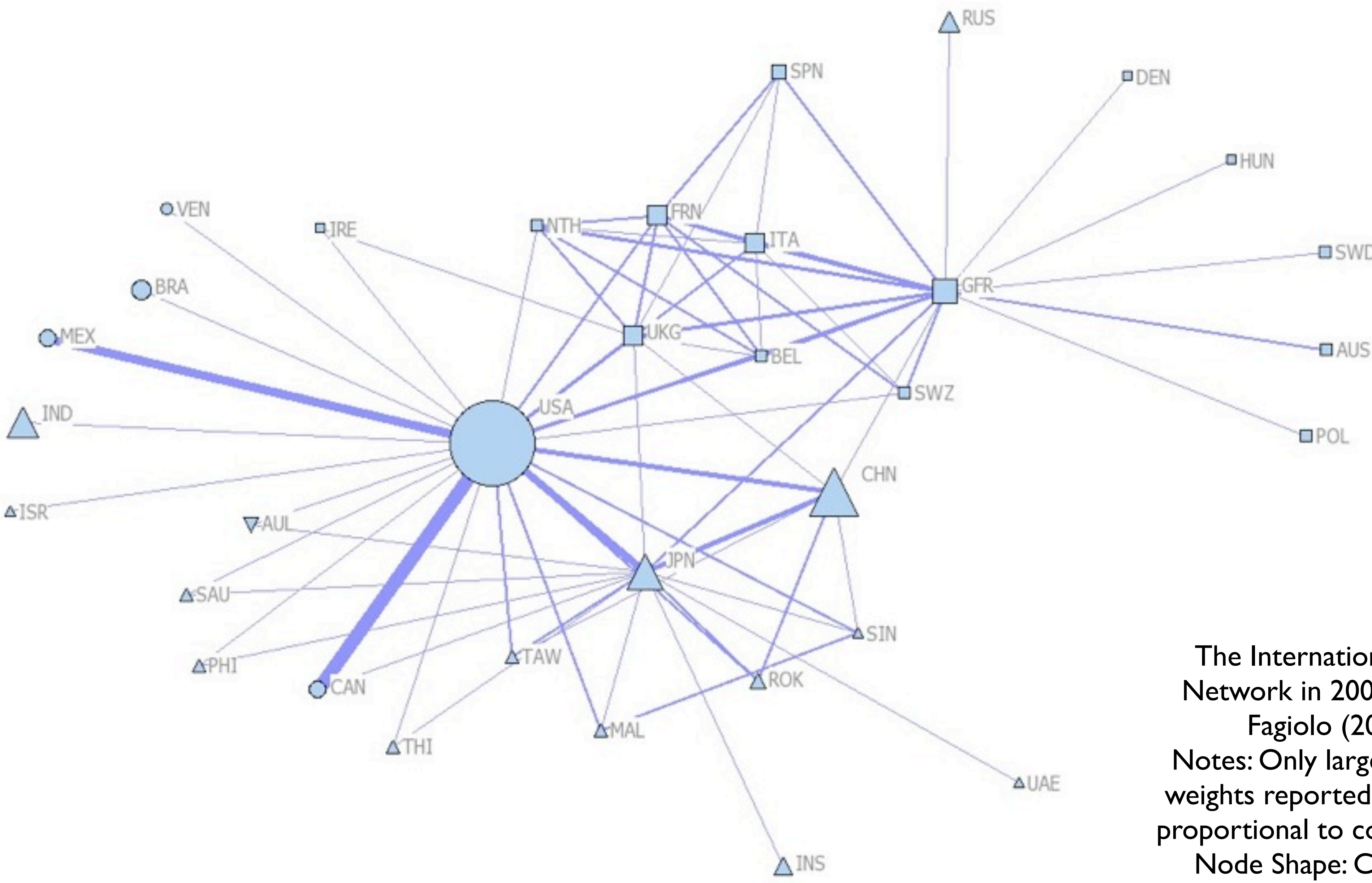
- Network data can be stored using the edge-list representation in a more efficient way especially if N is large and L is relatively small as compared to N(N-1)/2
- Network data can be complemented by: (i) node-specific variables; (ii) link attributes (other than weights)
- There are many available software packages that efficiently visualize network data:
  - ✓ Gephi (http://gephi.org): open source, very flexible and powerful
  - ✓ Netdraw (http://www.analytictech.com): less sophisticated but allows for a lot of computations
  - ✓ Visant (http://visant.bu.edu/): available as Java applet
  - ✓ Cytoscape (http://www.cytoscape.org): extremely powerful, especially for very large nets
  - ✓ Specific packages in Python (matplotlib) and R (igraph)

Netdraw: Manual Layout



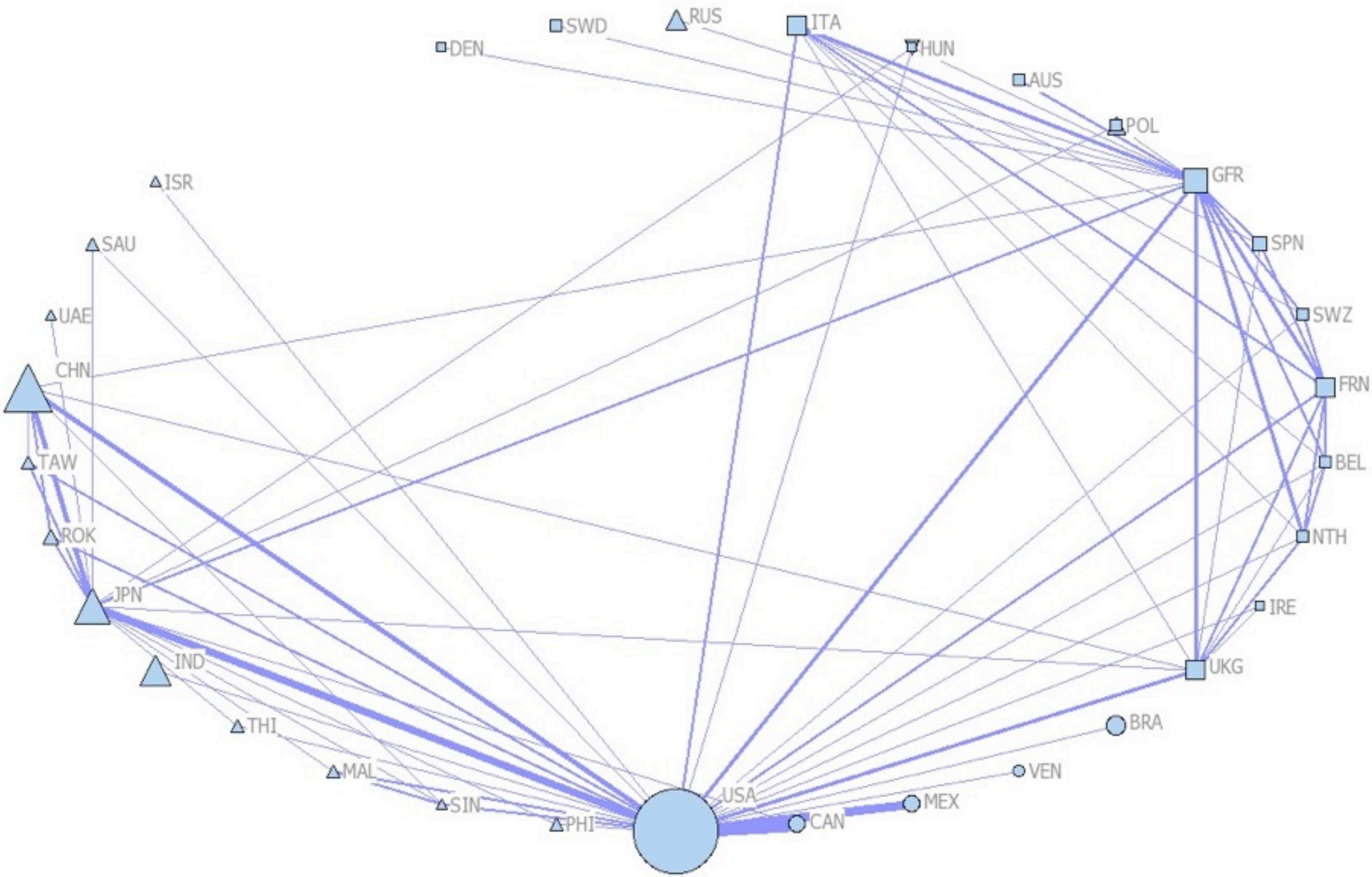
The International Trade Network in 2000. Source: Fagiolo (2010).
Notes: Only largest 1% link weights reported. Node size proportional to country GDP. Node Shape: Continent

Netdraw: Spring Embedding



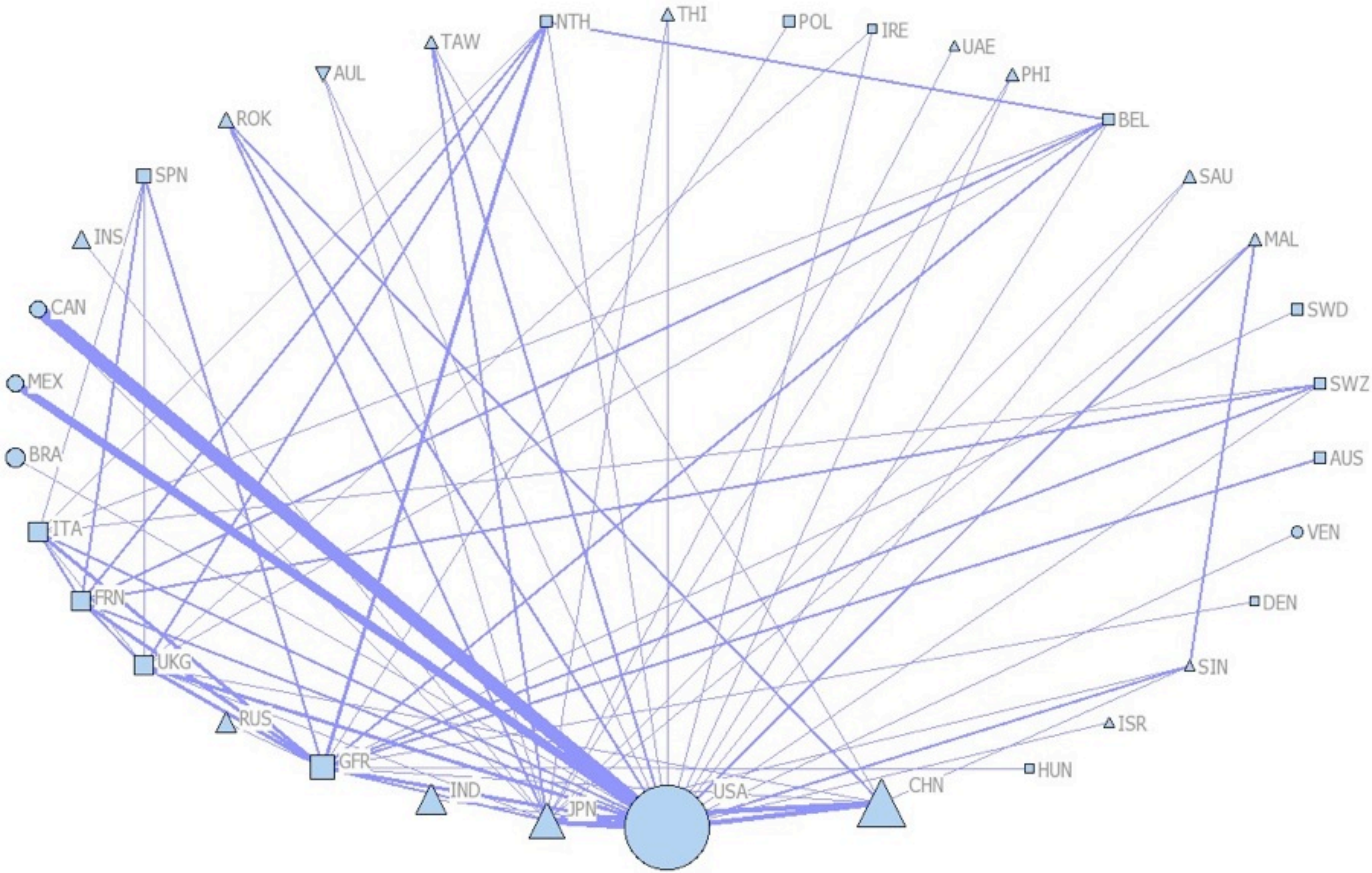
The International Trade Network in 2000. Source: Fagiolo (2010).
Notes: Only largest 1% link weights reported. Node size proportional to country GDP.
Node Shape: Continent

Netdraw: Circle (Labels)



The International Trade Network in 2000. Source: Fagiolo (2010).
Notes: Only largest 1% link weights reported. Node size proportional to country GDP. Node Shape: Continent

Netdraw: Circle (Per-Capita GDP)

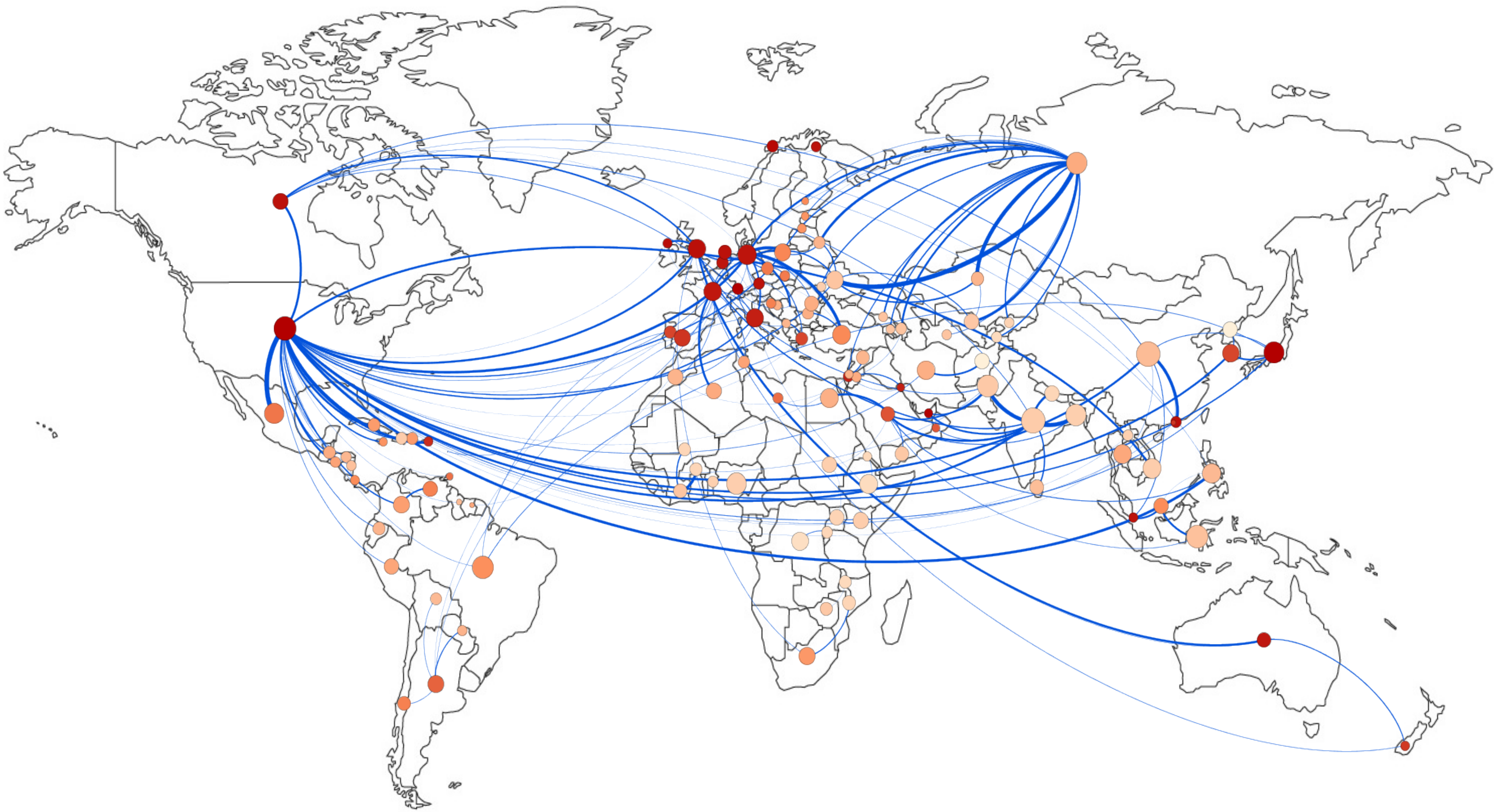


The International Trade Network in 2000. Source: Fagiolo (2010).
Notes: Only largest 1% link weights reported. Node size proportional to country GDP. Node Shape: Continent

Network Viz with Gephi

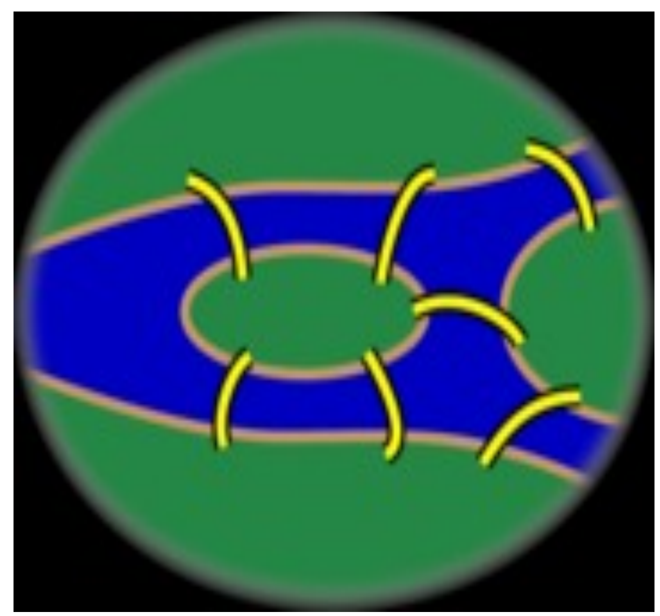
- Load IMN data: node and link attributes as csv tables
- Simplify viz using filters on link-weight levels
- Delete nodes that have a deg=0 after filtering
- Apply GEO layout
- Color links according to their weights
- Set node size according to population
- Set node color according to pcGDP
- Play with statistics and node partitioning
- Finish the job in Preview

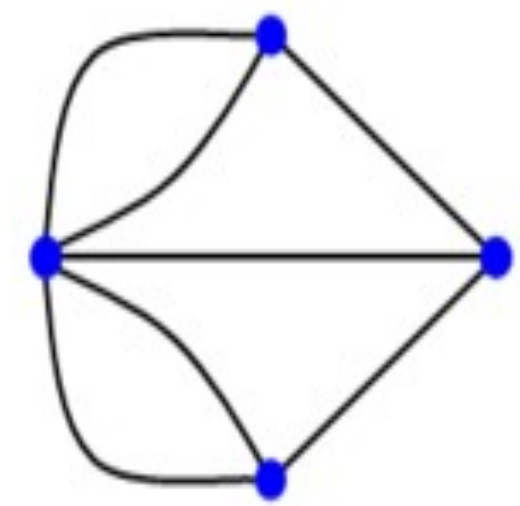
Network Viz with Gephi



Seven Bridges of Königsberg

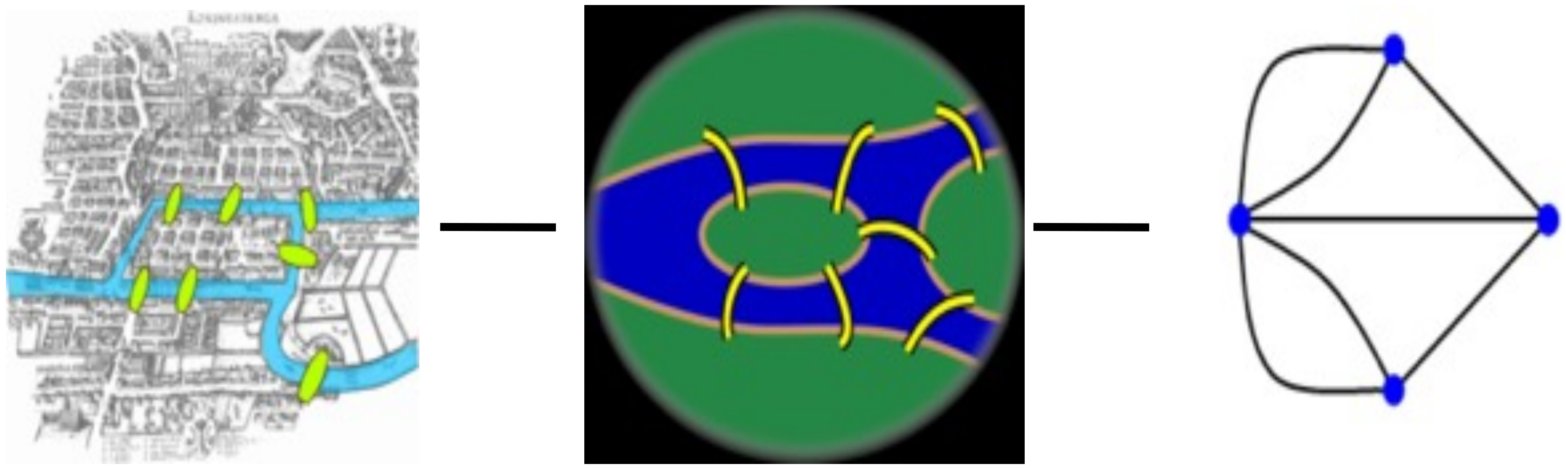
Problem of the Königsberg (now Kaliningrad, Russia) bridges: find a walk through the city that would cross each bridge once and only once





Euler proved that a solution does not exist and provided conditions for the general problem.

Seven Bridges of Königsberg



- Sketch of the proof
  - Transform the problem into a binary undirected graph: there are 4 nodes (areas) and 7 links (bridges)
  - In general, nodes can have odd or even degree. If a node has an even degree (>0) then the node is a crossing point: we can enter the node by one bridge and exit by another one (check for d=2 and generalize for d>2 even). If a node has an odd degree than it must be either a starting or an ending point of the path
  - One can pass over every bridge exactly once only if the number of vertices with odd degrees are either zero (starting and ending point coincide) or two (starting and ending points are different). Note: since the sum of all degrees must be even one cannot have a graph where there is just one node with odd degree.
  - Königsberg graph has 4 nodes, all with an odd degree!

What's Next

- What is a network? Examples of networks
- Why networks are important for economists?
- Networks and graphs
- Measures and metrics on networks
- Distributions of metrics and measures in large networks
- Modeling Networks
- Economic applications