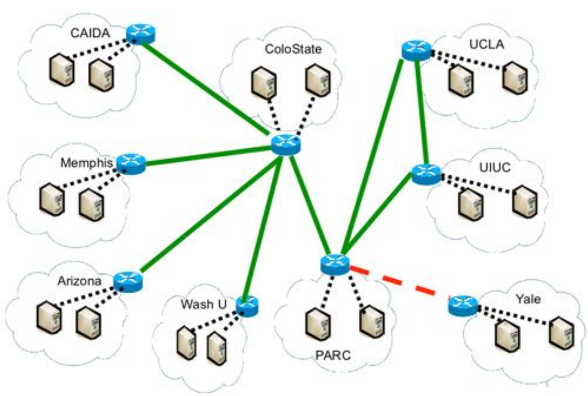
Complex networks: From Facebook to football

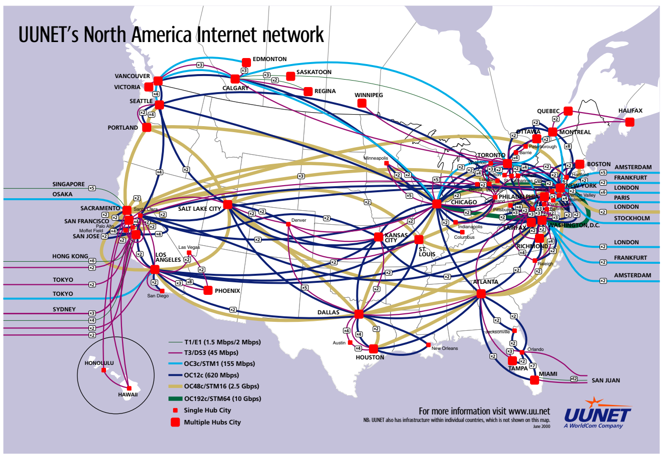
Hugo Touchette

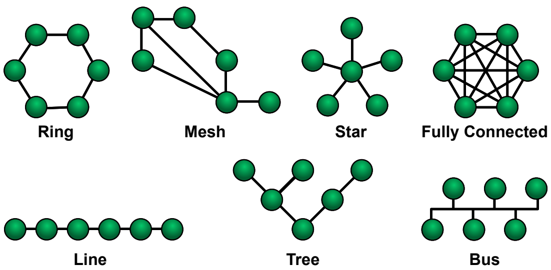
National Institute for Theoretical Physics (NITheP)
Stellenbosch

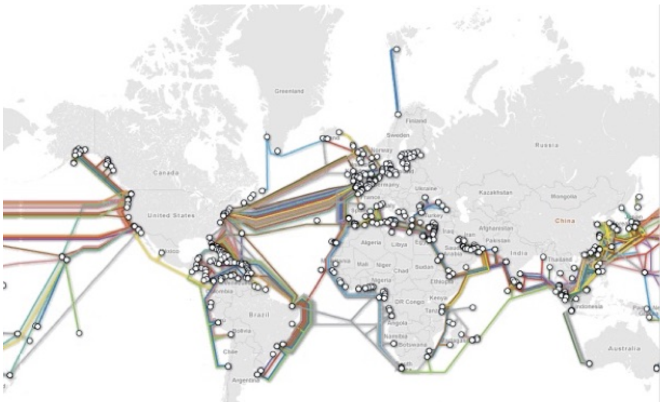
Department of Physics
Cape Town University
5 August 2015

Internet



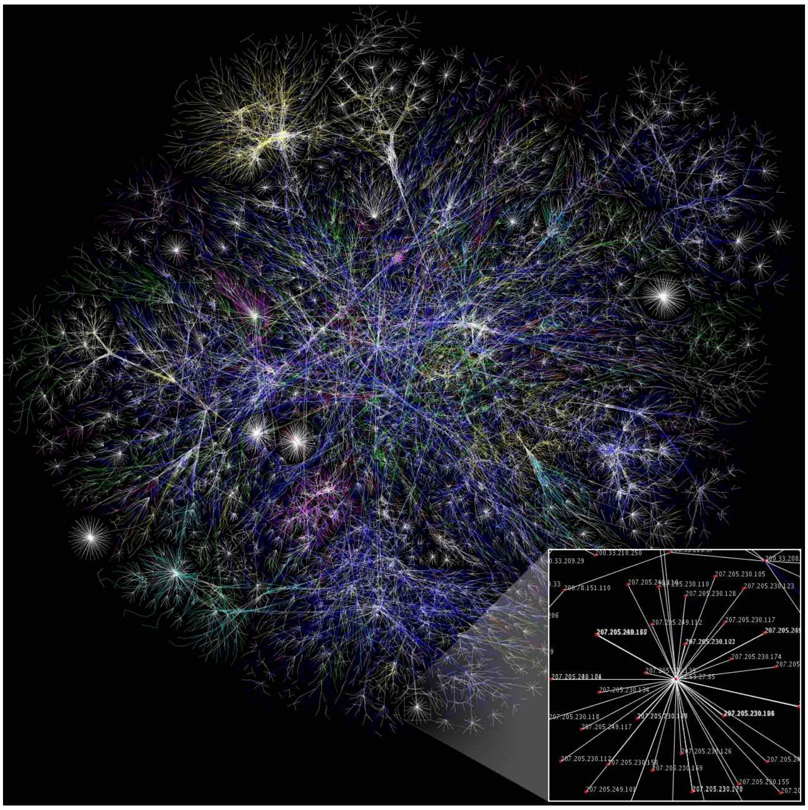




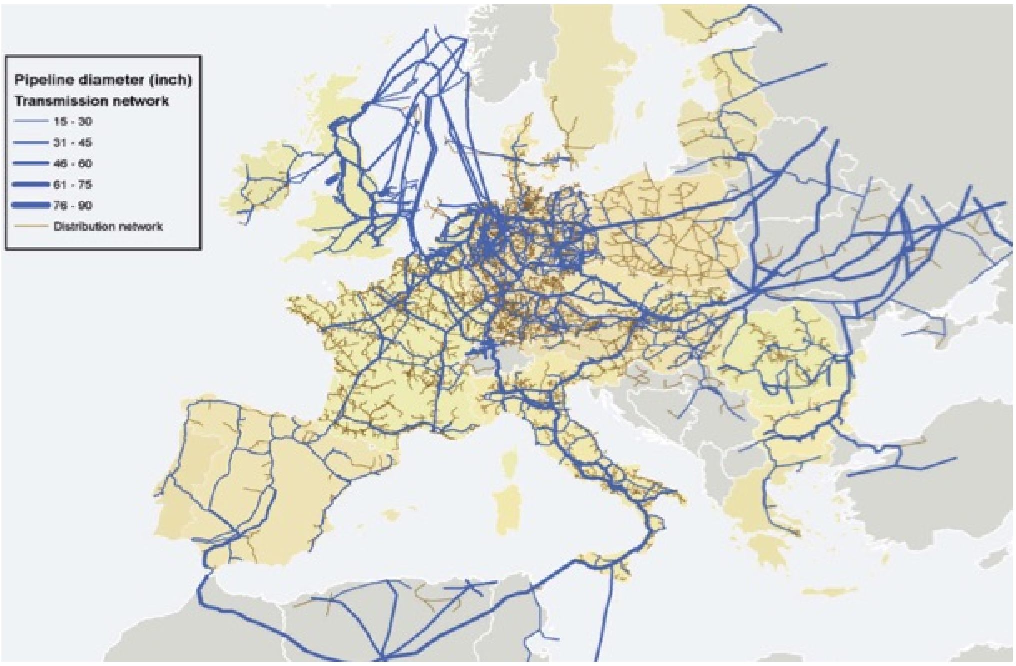


Internet (cont'd)

IP connection map, 15 Jan 2005 (from wiki)

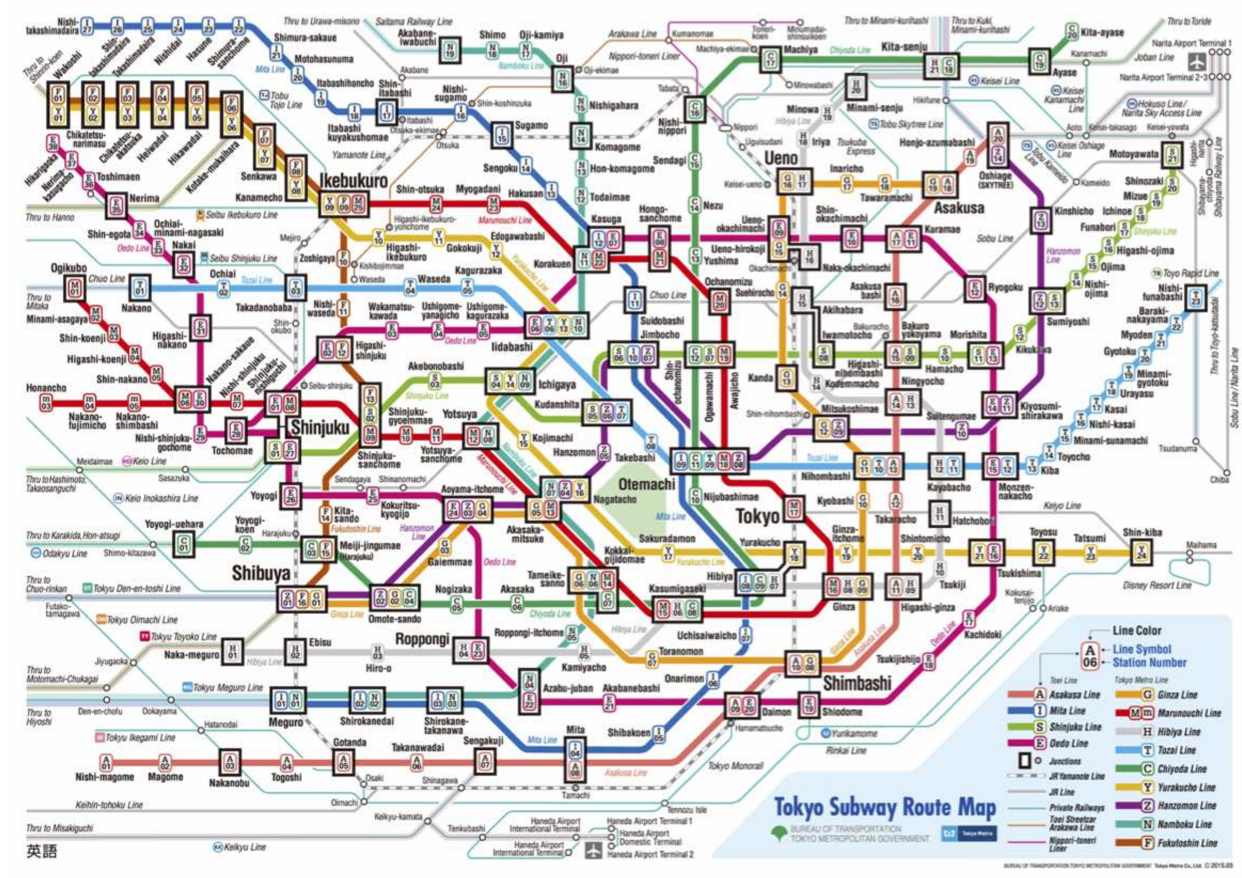


European gas network



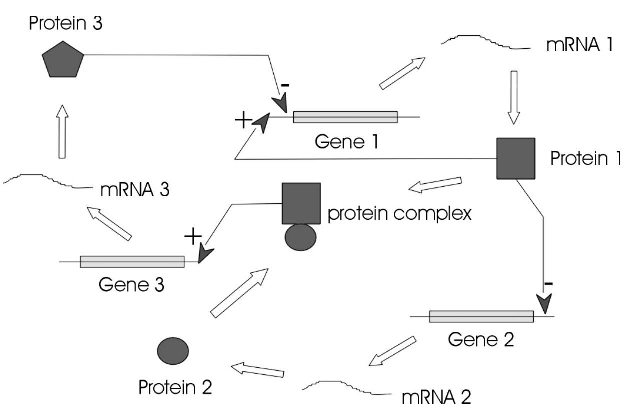
- $\sim$ 24 000 nodes, 25 000 links

Transportation

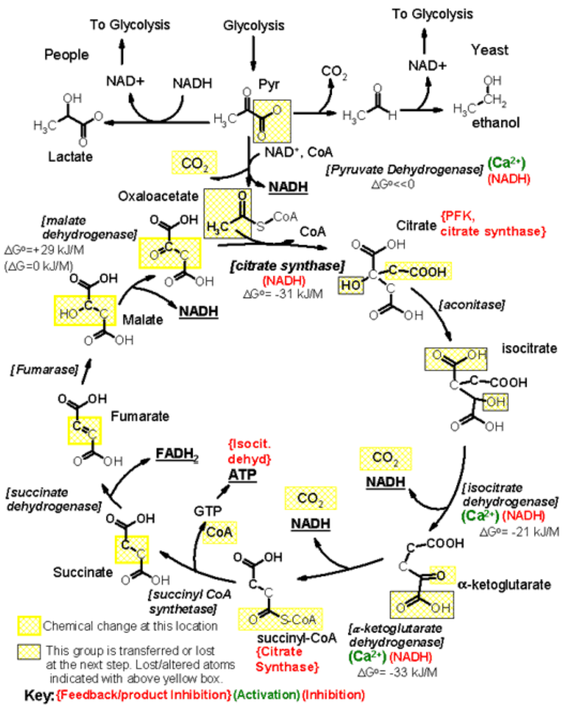


Biological and chemical networks

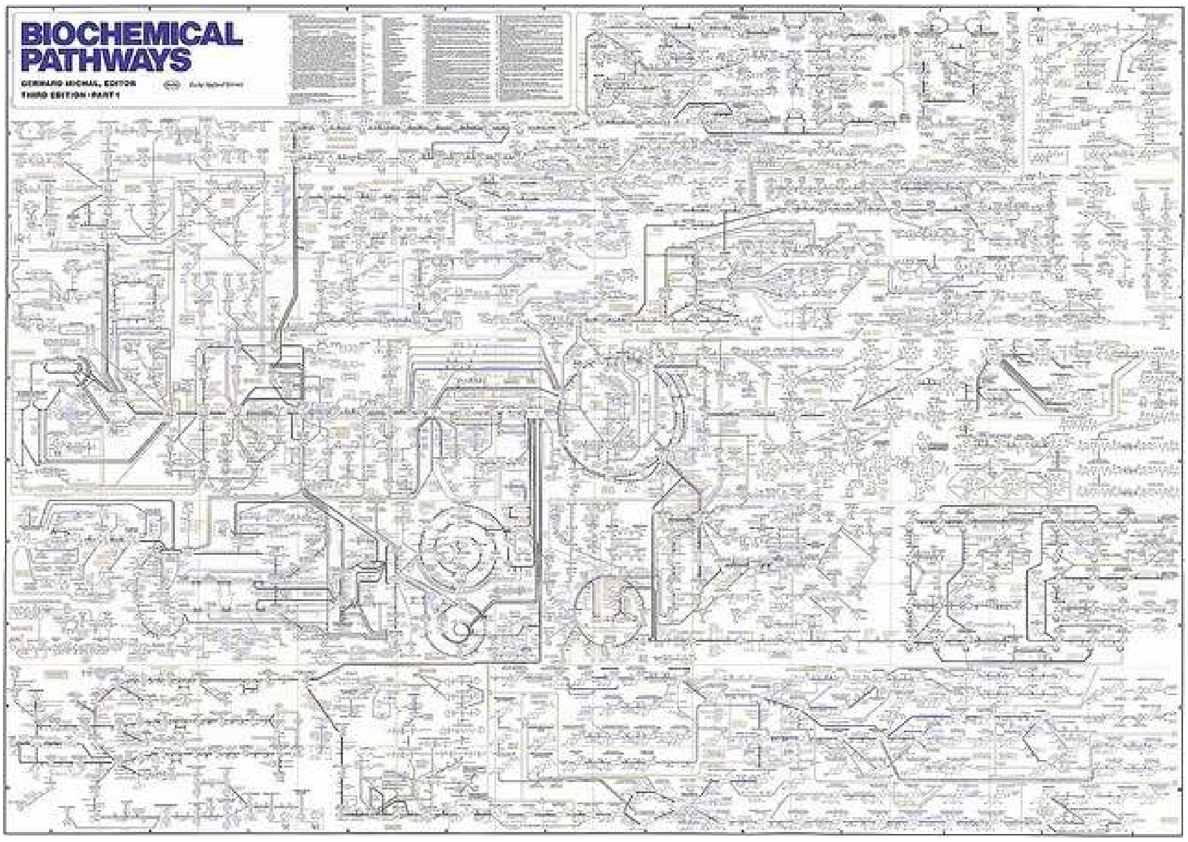
Gene regulatory network



Citric acid cycle

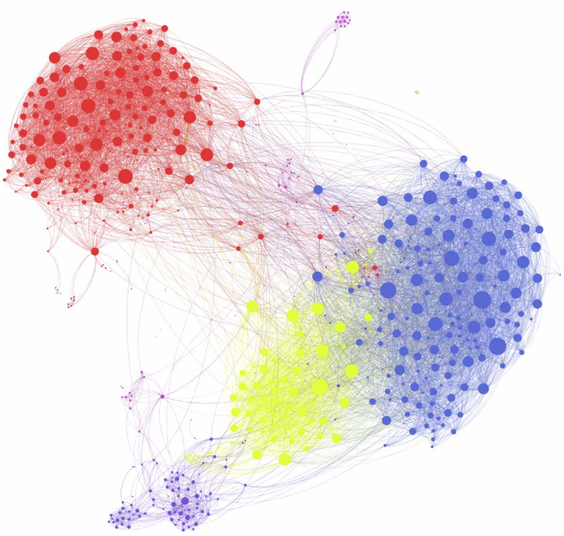


Biological and chemical networks (cont'd)

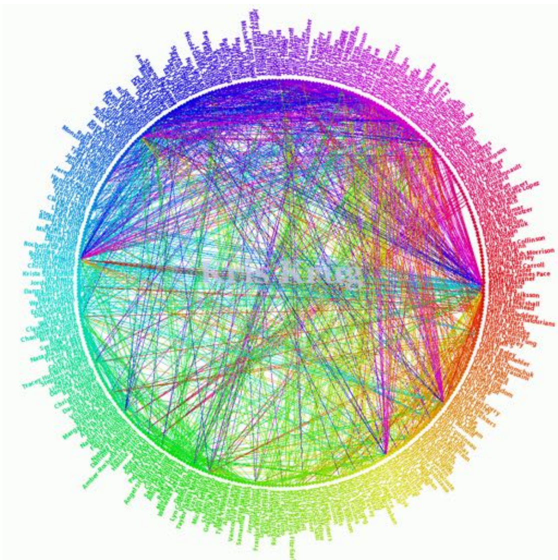


Facebook

Brendan Griffen



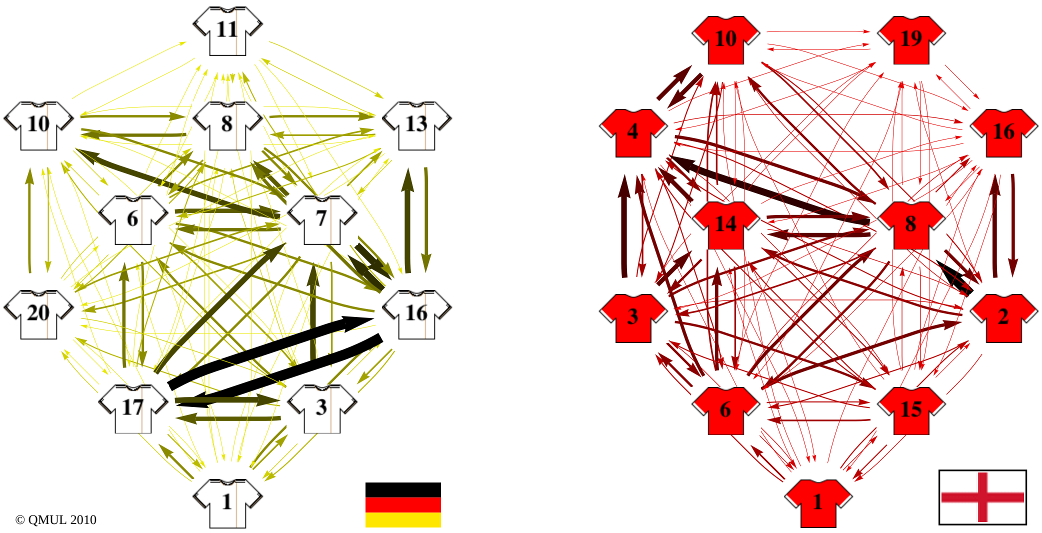
Kris Krüg



- Draw your Facebook network:
  https://msalganik.wordpress.com/2013/02/11/visualizing-your-facebook-network/
- More graphs:
  http://noduslabs.com/research/inclusive-exclusivity-innovative-networks/
  https://msalganik.wordpress.com/2013/02/24/a-gallery-of-personal-networks-from-facebook/

Football networks

2010 FIFA World Cup



Bits of graph theory

- $G = \{\text{nodes}, \text{edges}\}$
- Incidence/adjacency matrix:

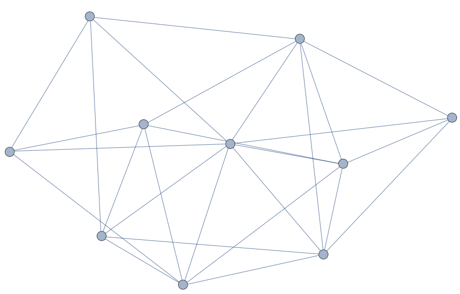
$$A_{ij} = \begin{cases} 1 & i,j \text{ linked} \\ 0 & \text{otherwise} \end{cases}$$

- Node degree:

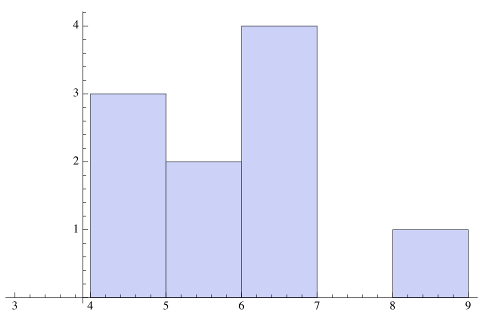
$$k_i = \sum_j A_{ij}$$

- Degree distribution: $P(k)$
- Shortest path (geodesic)
- Connected vs unconnected
- Largest component

10 nodes, 27 edges

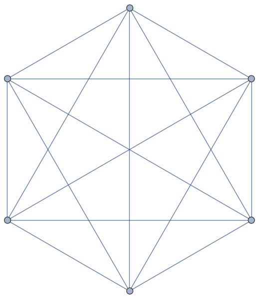


$\{4, 6, 6, 5, 4, 6, 4, 8, 5, 6\}$

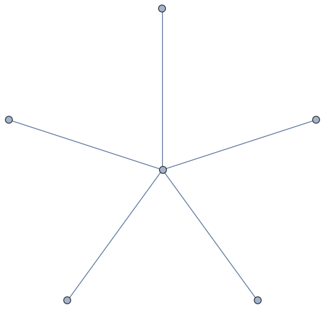


Regular graphs

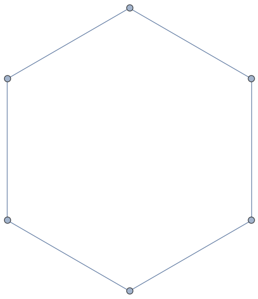
Complete



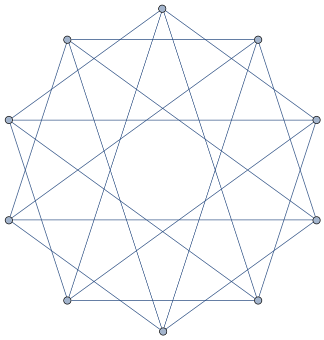
Star



Cycle



Circulant

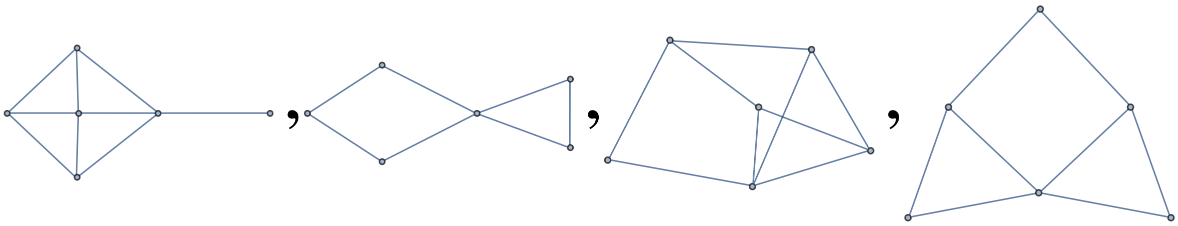


- Lattices
- ...

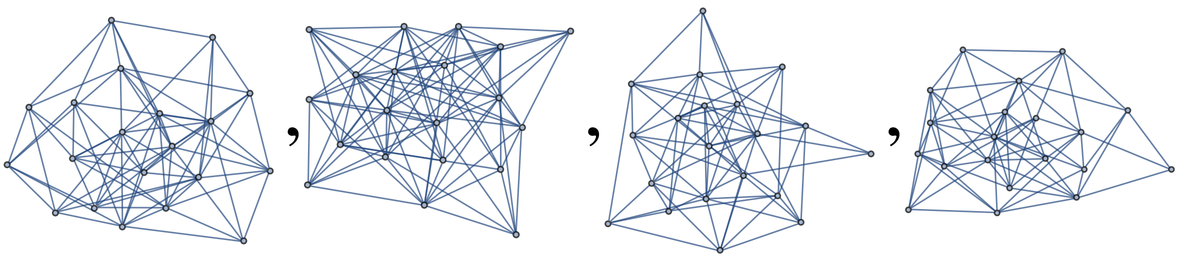
Random graphs

[Erdös & Rényi, 1959]

- N nodes
- Connect with probability p
- $(6, 0.5)$:



- $(20, 0.5)$:

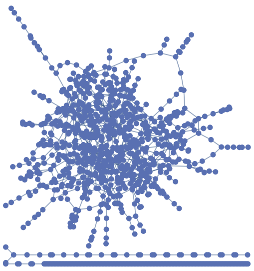
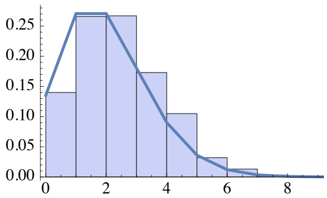


Random graphs (cont'd)

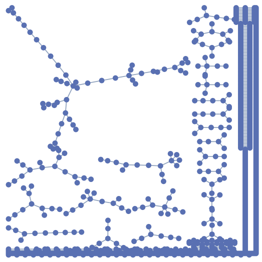
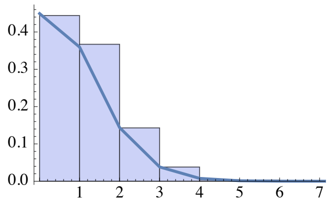
- Poisson degree distribution:

$$P(k) \to \frac{(Np)^k}{k!} e^{-Np}, \qquad N \to \infty, Np = \text{const}$$

- $Np < 1$: Likely disconnected
- $Np \geq 1$: Giant component
- Related to percolation transition

$N = 1000,\ Np = 2$

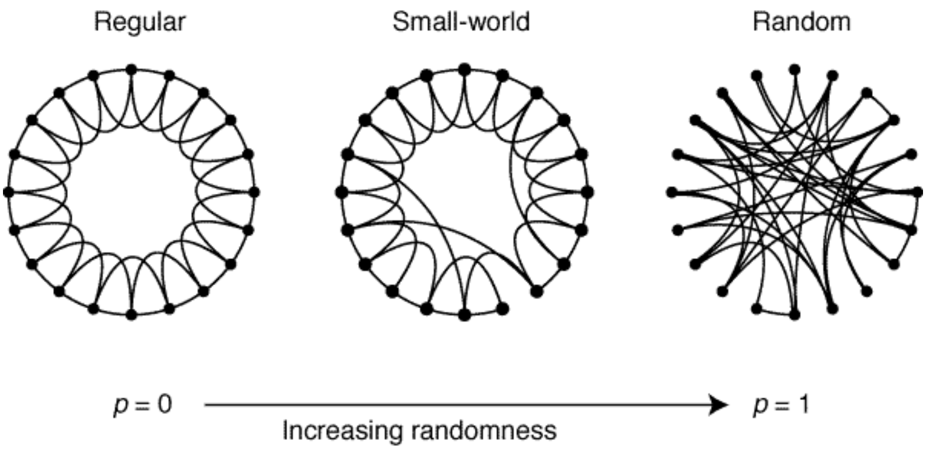



$N = 1000,\ Np = 0.8$

Small-world networks

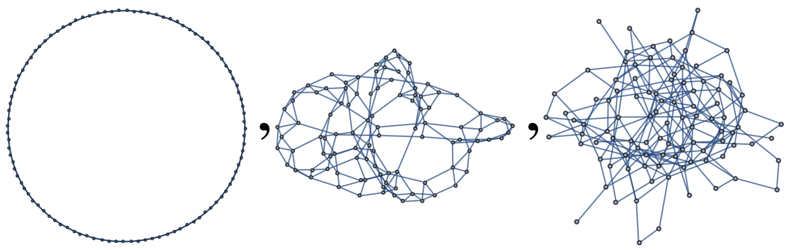
[Watts & Strogatz, 1998]

- Start with regular graph
- Rewire each edge at random with probability p
- Uniform re-attachment

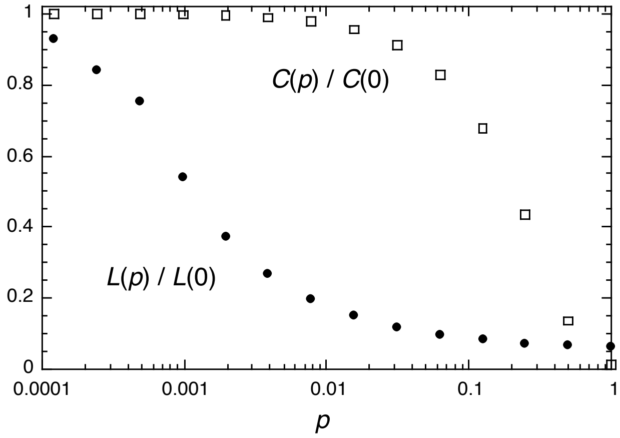


Small-world networks (cont'd)

- Efficient covering
- Small path length L
- High clustering C

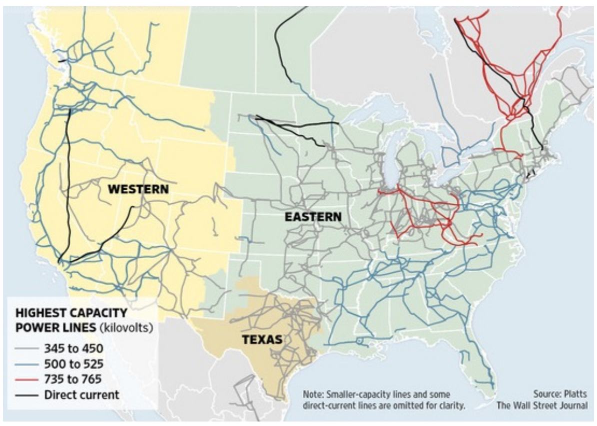


$N = 100,\ p = 0, 0.1, 1$



Applications

- Electricity grids
- Gene networks
- Social networks
- ...



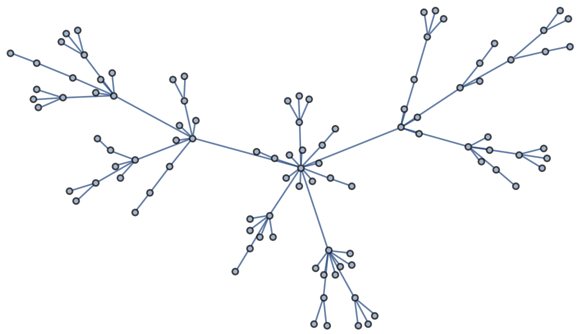
6 degrees of separation

- Kevin Bacon game (co-acting relation)
- No center: any actor/person will do
- Facebook app (average = 5.73)
- Erdös number
- Erdös-Bacon number (Sagan = 4 + 2, Feynman = 3 + 3)

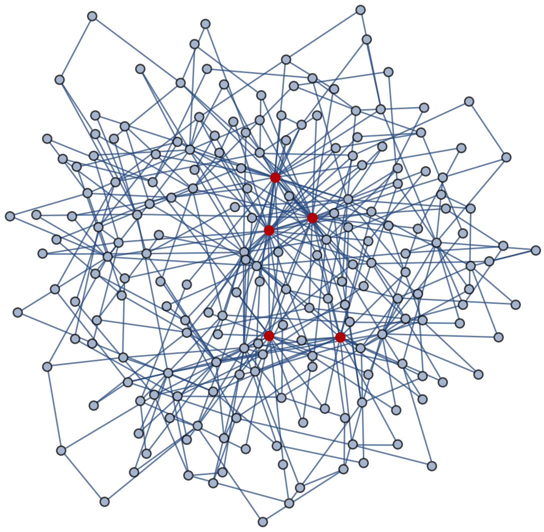
Scale-free networks

[Barabási & Albert, 1999]

- Start with n nodes
- Add new vertex with k edges
- Link edges to vertices at random in proportion to degree
- Preferential attachment



$N = 100, k = 1$



$N = 200, k = 2$

Scale-free networks (cont'd)

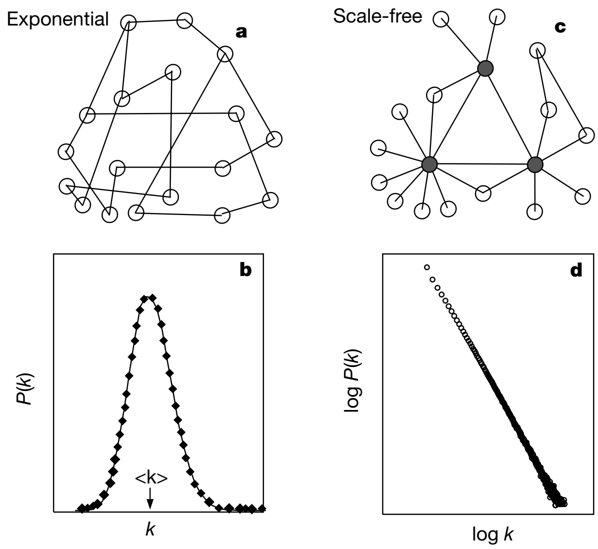
- Degree reinforcement: High degree nodes get more links
- Power-law degree distribution:

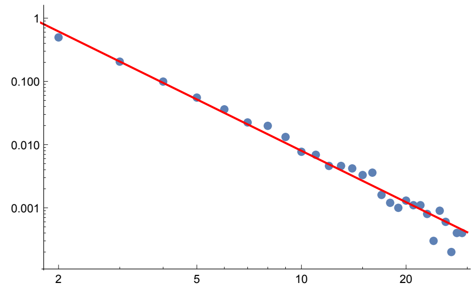
$$P(k) \sim \frac{1}{k^\gamma}, \qquad 2 < \gamma < 3$$

- Large hubs more likely

Examples

- Internet
- Scientific citations
- Neural networks
- Airline connections
- Stars?...

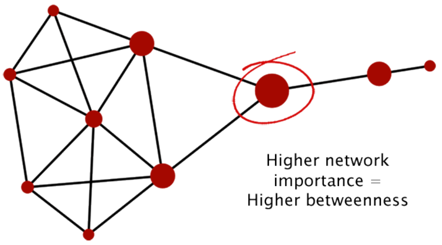


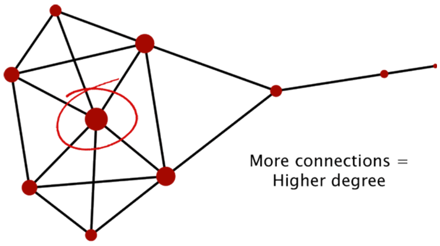


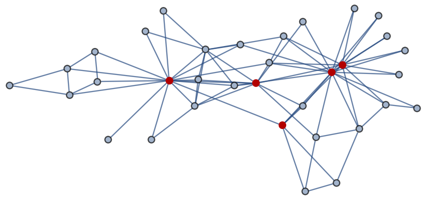
$N = 10^4, \ k = 2$

Hubs, centrality, and communities

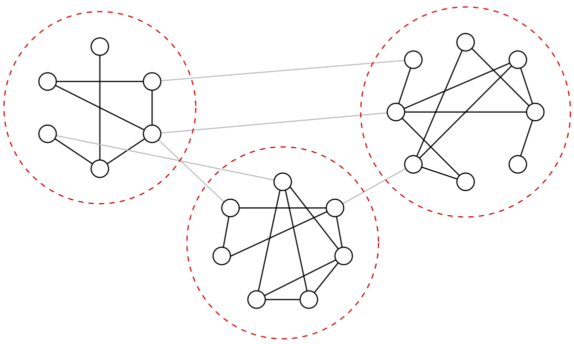
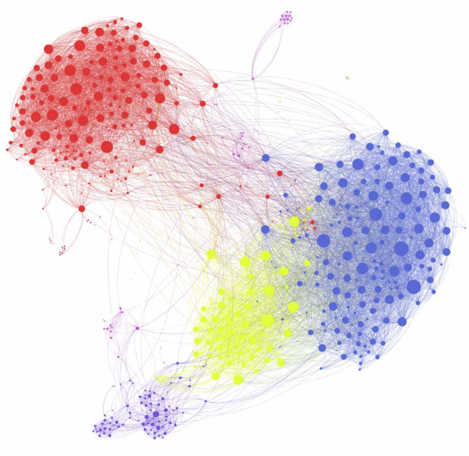
- **Hubs:** Central node, high centrality





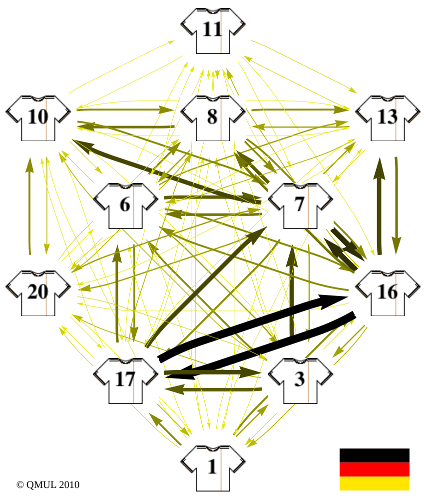


- Many definitions: Betweenness, degree centrality, pagerank, etc.
- **Communities:** Clusters, tight subgraphs

Football graphs

- With Javier López Peña
  (UCL, now at Kickdex London)
- 2010 FIFA World Cup in South Africa
- Passing data available (at the time)
- Data gathered with python script
- Pass networks generated with Mathematica
- Cumulative passes (all games added)
- Various centrality measures studied
- Predicted final result!
- www.maths.qmul.ac.uk/~ht/footballgraphs/
- Article: arxiv.org/abs/1206.6904



Measures

- Pass matrix: $A_{ij} = \#$ passes between players i, j
- Geodesic length: $d_{ij} =$ shortest (weighted) path between i, j

Closeness centrality

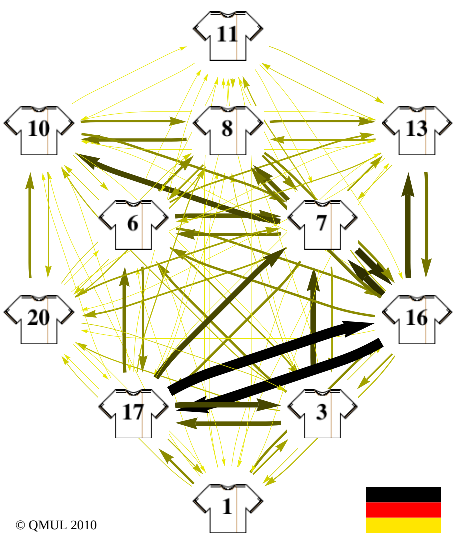
$$C_i = \frac{1}{\sum_{j \neq i} d_{ij} + \sum_{j \neq i} d_{ji}}$$

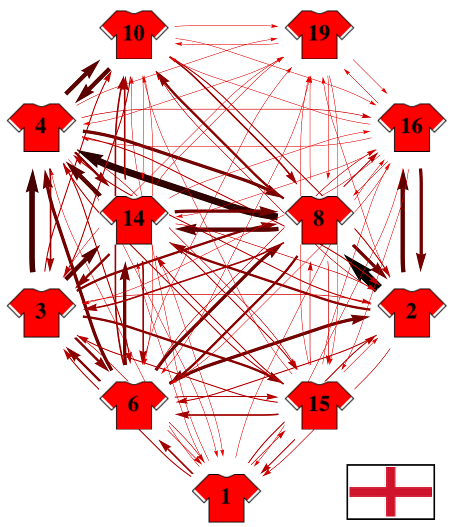
- High closeness = small distance
- Easily accessible player

Other quantities

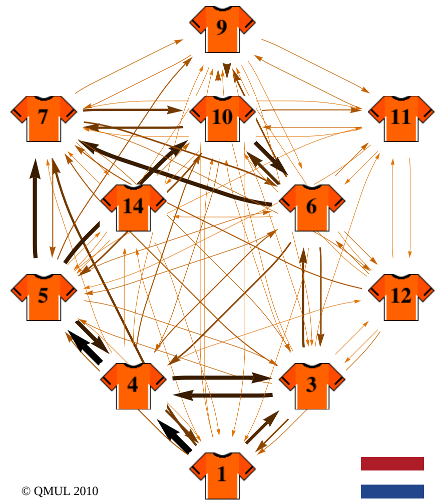
- Betweenness $C_B(i)$: Impact of removing a player (Hub related, 0 = player not involved)
- Pagerank x_i: Fraction of time player has the ball
- Clustering coefficient c_i^w: Clique factor

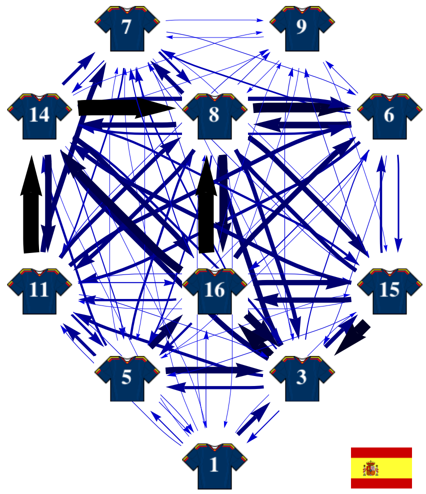
Germany vs England





The Netherlands vs Spain



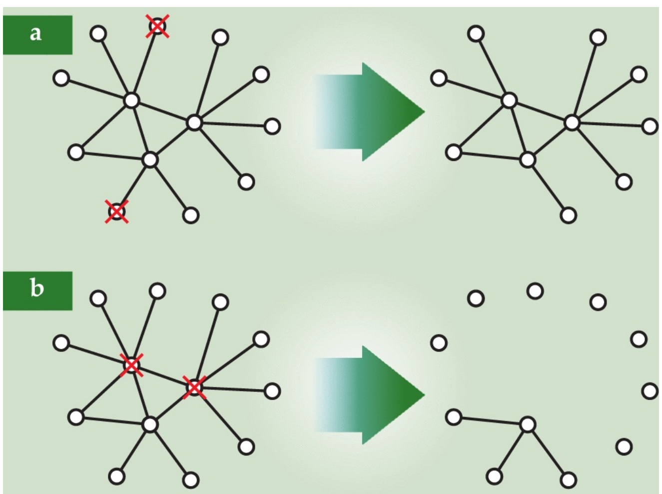


	C_i	$C_B(i)$	x_i	c_i^w
1 Stekelenburg	**16.34**	0.32	7.63	28.35
3 Heitinga	16.23	2.67	**11.06**	31.34
4 Mathijsen	**17.30**	1.30	10.84	33.22
5 V. Bronckhorst	15.74	1.12	10.07	**37.00**
6 Van Bommel	12.46	**3.08**	**11.19**	32.36
7 Kuyt	7.97	1.67	9.02	27.06
9 Van Persie	6.89	2.92	5.88	20.13
10 Sneijder	10.91	2.17	10.32	**33.77**
11 Robben	5.91	0.16	4.91	23.91

	C_i	$C_B(i)$	x_i	c_i^w
1 Casillas	16.52	0.00	3.29	20.46
3 Pique	17.32	**3.92**	11.46	30.70
5 Puyol	16.32	2.86	7.92	27.12
6 Iniesta	14.60	0.50	8.54	31.03
7 Villa	8.68	0.50	5.89	23.96
8 Xavi	**18.28**	1.19	**14.66**	**46.47**
11 Capdevila	16.54	**6.12**	10.56	29.91
14 Alonso	17.11	1.19	12.31	**41.69**
15 Ramos	16.45	2.41	9.02	27.05
16 Busquets	**18.55**	2.41	**12.99**	35.32

Research on complex networks

- Resilience/robustness of networks (random or targeted attacks)
- Immunology (centrality, communities)
- Layered networks (gas + electricity, distribution + consumption)
- Time-evolving networks (cities, transportation)
- Random walks on networks (Google, search, infections)



Useful references

D. J. Watts
Small Worlds
Princeton University Press, 1999

M. Newman
Networks
Oxford University Press, 2010

M. Newman
The physics of networks
Physics Today, Nov 2008

S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang
Complex networks: Structure and dynamics
Physics Reports 424, 175, 2006

R. Albert, A.-L. Barabási
Statistical mechanics of complex networks
Review of Modern Physics 74, 47, 2002