Network models

19th Feb 2014

Jesper Ferkinghoff-Borg
What is a network?

A network consists of *nodes* and *links*. Links are either *undirected* or *directed*.

**Examples (Nodes / Links):**

- Social Networks: Persons / friendships
- WWW: documents / URLs
- Economic Networks: Stocks / correlations
- Transport Networks: Airports / flights
- Biochemical networks:
  - Metabolic:
  - Transcriptional/regulatory:
  - Signaling:
- Protein interaction networks:
Network measures and concepts

- **Degree of a node**
  Number of links connected to the node
  For a directed graph, 2 types of degrees: *In-degrees* and *out-degrees*

- **Hub**
  A node with “high” (in/out) degree

- **Adjacency matrix**
  \[ A_{ij} = \begin{cases} 
  1 & \text{if link from } j \text{ to } i \\
  0 & \text{otherwise} 
  \end{cases} \]

- **Distance between two nodes = shortest path**
  Path length between two nodes defined by number of links in the path.
Network measures and concepts

- **Modularity**

A module is a region with *dense internal connections* and *sparse external connections*.

*Nodes in modules reflect a particular relationship with each other*

*modular nodes, peripheral nodes, interconnecting nodes, bridging nodes*

Modularity measures: Betweenness centrality and clustering coefficients
Network measures and concepts

- **Betweenness centrality**

  Measures the contribution of a node to the communication between all nodes

  \[ C_b(n) = \sum_{(s,t) \neq n} \frac{\rho_{st}(n)}{\rho_{st}} \]

  - number of shortest path between s and t passing through n
  - number of shortest path between s and t

  Small betweenness
  
  High betweenness
Network measures and concepts

- Clustering coefficient

\[ C(n) = \frac{2L_{\text{neighbours}}(n)}{d(n) \cdot (d(n) - 1)} \]

Number of links between neighbours of n

degree of node n

\[ C(A) = \frac{2}{20} \]

k = 5

[Diagram of a network with nodes and links, labeled with k = 5]
Random networks

- **Erdos-Renyi network**
  
  Start with N nodes and connect each pair of nodes with a probability $p$.

  How many links on average?

  - Degree distribution shows a *characteristic size* (*Poisson distribution*) and no modularity
  - Tail decreases *exponentially* (hubs are rare!)
  - Mean path length \( l \sim \log(N) \)
Scale free networks

- Degree distribution follows a power law

\[ P(k) \sim k^{-\gamma}, \quad 2 < \gamma < 3 \]

-(Approx.) Scaling behavior typical for many biological and non-biological networks.
- Properties of determined by a relatively small number of highly connected nodes (hubs)
- Basic scale free network models (preferential attachment) shows no modularity
- Average path length \( l \sim \log(\log(N)) \)

Degree distribution

Clustering coefficient
Hierarchical networks

- Combining modularity with scale-free-ness
- Hierarchical modularity:

\[ C(k) \sim k^{-1} \]

- Communication between modules maintained by few hubs

![Hierarchical network diagram](image_url)

- [Node distribution graph](image_url)
- [Cluster coefficient graph](image_url)

**Figure 3**

Network models are crucial for shaping our understanding of complex networks and help to explain the origin of observed network characteristics. There are three models that had a direct impact on our understanding of biological networks. The Erdös–Rényi (ER) model of a random network accounts for the coexistence of modularity, local clustering and scale-free topology in many real systems, it has to be assumed.
Networks in cellular systems

Networks constructed from:
- Biochemistry, genomics, microarray, yeast two hybrid, mass spectrometry, gene knockout / knockdown (RNAi), fluorescence imaging (GFP/FRET), text mining.

What are links and nodes in the different networks?
Networks in cellular systems

Y2H for protein-protein interaction

A. Regular transcription of the reporter gene

B. One fusion protein only (Gal4-BD + Bait) - no transcription

C. One fusion protein only (Gal4-AD + Prey) - no transcription

D. Two fusion proteins with interacting Bait and Prey
Networks in cellular systems

- Protein interaction networks in yeast (2H)

Hubs, connectivity and lethality.
Networks in cellular systems

- **Protein-protein Interaction networks** (> 7000 interactions)
  - MIPS (Munich Information center for Protein Sequence, interaction db.)
  - BIND (Biomolecular Interaction Network Database)
  - DIP (Database of interacting Proteins)
  - ...
  - Human Protein Reference Database (HPRD)
  - STRING (in-house)

- **Metabolic networks**
  - KEGG (Kyoto Encyclopedia of Genes and Genomes)
  - BIGG (Biochemical Genetc and Genomics knowledge base)

- **Regulatory networks**
  - UniPROBE, JASPAR (data from microarray)
  - Phospho.ELM, PhosphoSite, NetPhorest, CBS-databases
    (Human post-translational modifications)
Relevance of network measures

- **Modularity:**
  - Signatures of modularity omnipresent. Modules <-> functionality

- **Robustness / scale-freeness**
  - Scale-free network robust against accidental failures
  - 80% removal, remaining 20% still form connected cluster
  - Attack vulnerability: removal of a few key hubs leads to disintegration

- **Hubs**
  - Connectedness important for deletion phenotype
  - Evolutionary conservatism
  - Centrality seems related to essentiality

- **Small-world phenomena**
  - Perturbing activity of any node may effect the behaviour of the whole network
Relevance of network measures

- **Degree and betweenness centrality:**
  - 2 types of hubs: *Party hubs (modular)* and *date hubs (bridging)*
  - Bridging nodes relate to proteins *mediating co-ordination between functions*. (=> non-hub proteins may be important as well)
  - Distinction between party and date hubs *relates to protein domains*

- Date hubs less evolutionary constrained that party hubs
- Party hubs contain fewer linear motifs.
- => Date hubs important for *signaling integration*
Network medicine

We must take an integrative approach to understand relationship between genotype and phenotype

Network states may be just as important for disease predictions and correlations.
Network medicine

- Human has approx 25,000 genes
- 10% of genes have a known disease association

Do disease genes have quantifiable characteristics?

From a network perspective:

Are disease genes placed randomly on the “interactome”? 
Network medicine

Box 3 | **Hypotheses of network medicine**

Network medicine is based on a series of widely used (and often unspoken) hypotheses and organizing principles that link network structure to biological function and disease. Next, we summarize some of the most frequently used hypotheses, and they are discussed in more detail in the main text.

**Hubs**
Non-essential disease genes (representing most known disease genes) tend to avoid hubs and segregate at the functional periphery of the interactome. *In utero* essential genes tend to associated with hubs.

**Local hypothesis**
Proteins involved in the same disease have an increased tendency to interact with each other.

**Corollary of the local hypothesis**
Mutations in interacting proteins often lead to similar disease phenotypes.

**Disease module hypothesis**
Cellular components associated with a specific disease phenotype show a tendency to cluster in the same network neighbourhood.

**Network parsimony principle**
Causal molecular pathways often coincide with the shortest molecular paths between known disease-associated components.

**Shared components hypothesis**
Diseases that share disease-associated cellular components (genes, proteins, metabolites or microRNAs) show phenotypic similarity and comorbidity.
- In *utero* essential genes (red) tend to be in functional centres
- Non-essential disease genes (blue) tend to segregate at periphery
Emergence of a disease is a combinatorial problem
- Different defects and perturbations resulting in similar disease phenotype
- A monogenotype may yield several pathophenotypes (ex. sickle cell disease)
Network medicine

- Identifying and validating disease modules

**a** Interactome reconstruction

**b** Disease gene (seed) identification

**c** Disease module identification

**d** Pathway identification

**e** Validation/prediction

**Potential sources**
- OMIM
- GWA study
- Literature

**Disease 1 module**
- Disease 1 protein
- Disease 2 protein
- Overlapping protein

**Disease 2 module**
- Known disease 2 protein
- Predicted disease 2 protein

**Functional homogeneity**
- Gene ontology
- Tissue specificity
- Phenotypic similarity

**Dynamic homogeneity**
- Co-expression
- Genetic interactions
- Drug response

**Validation**
- Disease genes
- Disease pathways
- Drug targets

**Prediction**
Network medicine

- Identifying disease gene candidates

**Linkage method:** (local hypothesis)
Direct interaction partners of a disease protein assoc. with same disease phenotype

**Disease module-based method:** (full network neighbourhood)
Proteins in same module assoc. with same disease. Low-coverage of cellular interaction maps may require additional experimental efforts.

**Diffusion-based method:** (information of full network topology)
More robust way of identifying ‘module’/ disease gene candidates
Network medicine / Diseasome

- Diseasome network:
  Nodes are *diseases*, links are molecular relationships between these

  Explains why certain group of diseases arises together *(comorbidity)*

**Links:**
- *Shared genes:*
  on average twice as likely to develop a particular disease if it shares genes with primary disease. However, many cases of no comorbidity. High comorbidity observed if mutations affect *same domain* of a protein.

- *Shared metabolic pathway:*
  Comorbidity confirms metabolic disease networks constructed from metabolic pathways.

- *Comorbidity*
<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aldosteronism</td>
</tr>
<tr>
<td>2</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>3</td>
<td>Anaemia, congenital</td>
</tr>
<tr>
<td></td>
<td>deserythropoietic</td>
</tr>
<tr>
<td>4</td>
<td>Asthma</td>
</tr>
<tr>
<td>5</td>
<td>Ataxia-telangiectasia</td>
</tr>
<tr>
<td>6</td>
<td>Atherosclerosis</td>
</tr>
<tr>
<td>7</td>
<td>Blood group</td>
</tr>
<tr>
<td>8</td>
<td>Breast cancer</td>
</tr>
<tr>
<td>9</td>
<td>Cardiomyopathy</td>
</tr>
<tr>
<td>10</td>
<td>Cataract</td>
</tr>
<tr>
<td>11</td>
<td>Charcot–Marie–Tooth</td>
</tr>
<tr>
<td></td>
<td>disease</td>
</tr>
<tr>
<td>12</td>
<td>Colon cancer</td>
</tr>
<tr>
<td>13</td>
<td>Complement component</td>
</tr>
<tr>
<td></td>
<td>deficiency</td>
</tr>
<tr>
<td>14</td>
<td>Coronary artery disease</td>
</tr>
<tr>
<td>15</td>
<td>Coronary spasm</td>
</tr>
<tr>
<td>16</td>
<td>Deafness</td>
</tr>
<tr>
<td>17</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>18</td>
<td>Enolase-β deficiency</td>
</tr>
<tr>
<td>19</td>
<td>Epidermolysis bullosa</td>
</tr>
<tr>
<td>20</td>
<td>Epilepsy</td>
</tr>
<tr>
<td>21</td>
<td>Fanconi’s anaemia</td>
</tr>
<tr>
<td>22</td>
<td>Fatty liver</td>
</tr>
<tr>
<td>23</td>
<td>Gastric cancer</td>
</tr>
<tr>
<td>24</td>
<td>Gilbert’s syndrome</td>
</tr>
<tr>
<td>25</td>
<td>Glaucoma 1A</td>
</tr>
<tr>
<td>26</td>
<td>Goitre congenital</td>
</tr>
<tr>
<td>27</td>
<td>HARP syndrome</td>
</tr>
<tr>
<td>28</td>
<td>HELLP syndrome</td>
</tr>
<tr>
<td>29</td>
<td>Haemolytic anaemia</td>
</tr>
<tr>
<td>30</td>
<td>Hirschprung disease</td>
</tr>
<tr>
<td>31</td>
<td>Hyperbilirubinaemia</td>
</tr>
<tr>
<td>32</td>
<td>Hypertension</td>
</tr>
<tr>
<td>33</td>
<td>Hypertension diastolic</td>
</tr>
<tr>
<td>34</td>
<td>Hyperthyroidism</td>
</tr>
<tr>
<td>35</td>
<td>Hypoaldosteronism</td>
</tr>
<tr>
<td>36</td>
<td>Leigh syndrome</td>
</tr>
<tr>
<td>37</td>
<td>Leukaemia</td>
</tr>
<tr>
<td>38</td>
<td>Low renin hypertension</td>
</tr>
<tr>
<td>39</td>
<td>Lymphoma</td>
</tr>
<tr>
<td>40</td>
<td>Mental retardation</td>
</tr>
<tr>
<td>41</td>
<td>Muscular dystrophy</td>
</tr>
<tr>
<td>42</td>
<td>Myocardial infarction</td>
</tr>
<tr>
<td>43</td>
<td>Myopathy</td>
</tr>
<tr>
<td>44</td>
<td>Nucleoside phosphorylase</td>
</tr>
<tr>
<td></td>
<td>deficiency</td>
</tr>
<tr>
<td>45</td>
<td>Obesity</td>
</tr>
<tr>
<td>46</td>
<td>Paraganglioma</td>
</tr>
<tr>
<td>47</td>
<td>Parkinson’s disease</td>
</tr>
<tr>
<td>48</td>
<td>Pheochromocytoma</td>
</tr>
<tr>
<td>49</td>
<td>Prostate cancer</td>
</tr>
<tr>
<td>50</td>
<td>Pseudohypoaldosteronism</td>
</tr>
<tr>
<td>51</td>
<td>Retinitis pigmentosa</td>
</tr>
<tr>
<td>52</td>
<td>Schizoaffective disorder</td>
</tr>
<tr>
<td>53</td>
<td>Spherocytosis</td>
</tr>
<tr>
<td>54</td>
<td>Spina bifida</td>
</tr>
<tr>
<td>55</td>
<td>Spinocerebellar ataxia</td>
</tr>
<tr>
<td>56</td>
<td>Stroke</td>
</tr>
<tr>
<td>57</td>
<td>Thyroid carcinoma</td>
</tr>
<tr>
<td>58</td>
<td>Total iodide organification</td>
</tr>
<tr>
<td></td>
<td>defect</td>
</tr>
<tr>
<td>59</td>
<td>Trifunctional protein</td>
</tr>
<tr>
<td></td>
<td>deficiency</td>
</tr>
<tr>
<td>60</td>
<td>Unipolar depression</td>
</tr>
</tbody>
</table>
Human disease network

Giant cluster
Human disease network

Shared genes link between breast and bone cancer
Human disease network

Comorbidity analysis

<table>
<thead>
<tr>
<th></th>
<th>Relative risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.15 ± 0.05</td>
</tr>
<tr>
<td>Gene sharing</td>
<td>0.25 ± 0.05</td>
</tr>
<tr>
<td>Domain sharing</td>
<td>0.35 ± 0.05</td>
</tr>
<tr>
<td>Gene sharing (no domain)</td>
<td>0.25 ± 0.05</td>
</tr>
</tbody>
</table>
Treatment strategies

- Understand and minimize side-effects
- Combining drugs
- Time-staggering
- Multiple targets (AIDS, cancer, depression)
- Palliative drugs (drugs targeting neighbours of disease protein)
- Personalized medicine
Conclusions

- Network models provide an convenient framework for cellular systems biology
- Topological properties can in many cases be linked to biological properties
- Network models emphasize the global / integrative functioning of the cell
- Construction of networks is currently challenged by:
  - lack of systematic, standardised, unbiased approaches
  - lack of rigorous mathematical methods for integration and validation
  - Incompleteness of data
- Networks are static (time and spatial averaging)
  - Real biological processes are dynamic (Network dynamics)
- Networks treat links and nodes uniformly
  - Not appropriate for detailed mechanistic understanding.
  - The rules of information flow still poorly understood
- 21 century: Complex systems modelling and hypothesis will be driven by 
  \textit{large scale multivariate and multi source data}. 