Organized Complexity and Network Science: Making Sense of Network Data

David Alderson, PhD
Operations Research Department
Naval Postgraduate School

dlalders@nps.edu

Graph Exploitation Workshop
MIT Lincoln Labs
April 2010
Acknowledgments

• Caltech: John Doyle, Lun Li, Steven Low
• AT&T-Research: Walter Willinger
• UCSB: Jean Carlson
• NPS Faculty: Bradley, Brown, Carlyle, Craparo, Dell, Denning, Eagle, Salmerón, Wood
• NPS Students: Barkley, Derosier, Langford, McPherson, Nicholas, Sanchez

• Supported by ONR MURI Grant: “Next-Generation Network Science” (Award #N0001408WR20242)
The Network Science “Sales Pitch”

• advances in info tech... massive data sets
• simply displaying them yields nonsensical results

http://research.lumeta.com/ches/map/

• claim: need new techniques to discover key structure
• claim: deep understanding of this structure will help predict (and perhaps control) system behavior
The Network Science “Sales Pitch”

• claim: networks are fundamental and universal
• the term “network” is ambiguous
  – a Rorschach test
  – a discrete approximation to any continuous relationship
• important: how to formalize the problem?
  – affects how we interpret data
  – can dictate the answers that we obtain

What is essential to our understanding of the structure (and function) of networks?

• This talk: revisiting “organized complexity”
  – function, constraints, tradeoffs, design
“mainstream network science”

- graphs are well-defined: $G(V,E)$
- structure, function, architecture = connectivity

- null hypothesis: random graphs
  - but real networks are not (uniformly) random!
  - connectivity is skewed, clustered, correlated

- develop novel descriptive models
  - large data samples, uncertainty
  - analysis of statistical properties
  - emphasis: “likely” configurations that match statistics
  - inference, prediction based on the properties of the model
  - statistical generalization: generating ensembles

<table>
<thead>
<tr>
<th>network</th>
<th>type</th>
<th>( n )</th>
<th>( m )</th>
<th>( z )</th>
<th>( \ell )</th>
<th>( \alpha )</th>
<th>( C^{(1)} )</th>
<th>( C^{(2)} )</th>
<th>( r )</th>
<th>Ref(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>film actors</td>
<td>undirected</td>
<td>449913</td>
<td>25516482</td>
<td>113.43</td>
<td>3.48</td>
<td>2.3</td>
<td>0.20</td>
<td>0.78</td>
<td>0.208</td>
<td>29, 416</td>
</tr>
<tr>
<td>company directors</td>
<td>undirected</td>
<td>7673</td>
<td>55392</td>
<td>14.44</td>
<td>4.60</td>
<td>–</td>
<td>0.59</td>
<td>0.88</td>
<td>0.276</td>
<td>105, 323</td>
</tr>
<tr>
<td>math coauthorship</td>
<td>undirected</td>
<td>253339</td>
<td>496489</td>
<td>3.92</td>
<td>7.57</td>
<td>–</td>
<td>0.15</td>
<td>0.34</td>
<td>0.120</td>
<td>107, 182</td>
</tr>
<tr>
<td>physics coauthorship</td>
<td>undirected</td>
<td>52909</td>
<td>245300</td>
<td>9.27</td>
<td>6.19</td>
<td>–</td>
<td>0.45</td>
<td>0.56</td>
<td>0.363</td>
<td>311, 313</td>
</tr>
<tr>
<td>biology coauthorship</td>
<td>undirected</td>
<td>1520251</td>
<td>11803064</td>
<td>15.53</td>
<td>4.92</td>
<td>–</td>
<td>0.088</td>
<td>0.60</td>
<td>0.127</td>
<td>311, 313</td>
</tr>
<tr>
<td>telephone call graph</td>
<td>undirected</td>
<td>47000000</td>
<td>80000000</td>
<td>3.16</td>
<td>2.1</td>
<td>–</td>
<td>0.16</td>
<td>0.16</td>
<td>0.92</td>
<td>31, 5, 49</td>
</tr>
<tr>
<td>email messages</td>
<td>directed</td>
<td>59912</td>
<td>86300</td>
<td>1.44</td>
<td>4.95</td>
<td>1.5/2.0</td>
<td>–</td>
<td>0.015</td>
<td>0.16</td>
<td>136</td>
</tr>
<tr>
<td>email address books</td>
<td>directed</td>
<td>16881</td>
<td>57029</td>
<td>3.38</td>
<td>5.22</td>
<td>–</td>
<td>0.17</td>
<td>0.13</td>
<td>0.092</td>
<td>321</td>
</tr>
<tr>
<td>student relationships</td>
<td>undirected</td>
<td>573</td>
<td>477</td>
<td>1.66</td>
<td>16.01</td>
<td>–</td>
<td>0.995</td>
<td>0.001</td>
<td>-0.029</td>
<td>45</td>
</tr>
<tr>
<td>sexual contacts</td>
<td>undirected</td>
<td>2810</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>265, 266</td>
</tr>
<tr>
<td>WWW nd.edu</td>
<td>directed</td>
<td>269504</td>
<td>1497135</td>
<td>5.55</td>
<td>11.27</td>
<td>2.1/2.4</td>
<td>0.11</td>
<td>0.29</td>
<td>-0.067</td>
<td>14, 34</td>
</tr>
<tr>
<td>WWW Altavista</td>
<td>directed</td>
<td>203549046</td>
<td>2130000000</td>
<td>10.46</td>
<td>16.18</td>
<td>2.1/2.7</td>
<td>0.13</td>
<td>0.15</td>
<td>0.157</td>
<td>244</td>
</tr>
<tr>
<td>citation network</td>
<td>directed</td>
<td>783339</td>
<td>6716198</td>
<td>8.57</td>
<td>3.0/–</td>
<td>0.066</td>
<td></td>
<td></td>
<td></td>
<td>351</td>
</tr>
<tr>
<td>Roget’s Thesaurus</td>
<td>directed</td>
<td>10222</td>
<td>5103</td>
<td>4.99</td>
<td>4.87</td>
<td>–</td>
<td>0.13</td>
<td>0.15</td>
<td>0.157</td>
<td>244</td>
</tr>
<tr>
<td>word co-occurrence</td>
<td>undirected</td>
<td>460902</td>
<td>17000000</td>
<td>70.13</td>
<td>2.7</td>
<td>–</td>
<td>0.44</td>
<td>0.119</td>
<td>0.157</td>
<td>119, 157</td>
</tr>
<tr>
<td>Internet</td>
<td>undirected</td>
<td>10697</td>
<td>31992</td>
<td>5.98</td>
<td>3.31</td>
<td>2.5</td>
<td>0.035</td>
<td>0.39</td>
<td>-0.189</td>
<td>86, 148</td>
</tr>
<tr>
<td>power grid</td>
<td>undirected</td>
<td>4941</td>
<td>6594</td>
<td>2.67</td>
<td>18.99</td>
<td>–</td>
<td>0.10</td>
<td>0.080</td>
<td>-0.003</td>
<td>416</td>
</tr>
<tr>
<td>train routes</td>
<td>undirected</td>
<td>587</td>
<td>19603</td>
<td>66.79</td>
<td>2.16</td>
<td>–</td>
<td>0.69</td>
<td>-0.033</td>
<td></td>
<td>366</td>
</tr>
<tr>
<td>software packages</td>
<td>directed</td>
<td>1439</td>
<td>1723</td>
<td>1.20</td>
<td>2.42</td>
<td>1.6/1.4</td>
<td>0.070</td>
<td>0.082</td>
<td>-0.016</td>
<td>318</td>
</tr>
<tr>
<td>software classes</td>
<td>directed</td>
<td>1377</td>
<td>2213</td>
<td>1.61</td>
<td>1.51</td>
<td>–</td>
<td>0.033</td>
<td>0.012</td>
<td>-0.119</td>
<td>395</td>
</tr>
<tr>
<td>electronic circuits</td>
<td>undirected</td>
<td>24097</td>
<td>53248</td>
<td>4.34</td>
<td>11.05</td>
<td>3.0</td>
<td>0.010</td>
<td>0.030</td>
<td>-0.154</td>
<td>155</td>
</tr>
<tr>
<td>peer-to-peer network</td>
<td>undirected</td>
<td>880</td>
<td>1296</td>
<td>1.47</td>
<td>4.28</td>
<td>2.1</td>
<td>0.012</td>
<td>0.011</td>
<td>-0.366</td>
<td>6, 354</td>
</tr>
<tr>
<td>metabolic network</td>
<td>undirected</td>
<td>765</td>
<td>3686</td>
<td>9.64</td>
<td>2.56</td>
<td>2.2</td>
<td>0.090</td>
<td>0.67</td>
<td>-0.240</td>
<td>214</td>
</tr>
<tr>
<td>protein interactions</td>
<td>undirected</td>
<td>2115</td>
<td>2240</td>
<td>2.12</td>
<td>6.80</td>
<td>2.4</td>
<td>0.072</td>
<td>0.071</td>
<td>-0.156</td>
<td>212</td>
</tr>
<tr>
<td>marine food web</td>
<td>directed</td>
<td>135</td>
<td>598</td>
<td>4.43</td>
<td>2.05</td>
<td>–</td>
<td>0.16</td>
<td>0.23</td>
<td>-0.263</td>
<td>204</td>
</tr>
<tr>
<td>freshwater food web</td>
<td>directed</td>
<td>92</td>
<td>997</td>
<td>10.84</td>
<td>1.90</td>
<td>–</td>
<td>0.20</td>
<td>0.087</td>
<td>-0.326</td>
<td>272</td>
</tr>
<tr>
<td>neural network</td>
<td>directed</td>
<td>307</td>
<td>2359</td>
<td>7.68</td>
<td>3.97</td>
<td>–</td>
<td>0.18</td>
<td>0.28</td>
<td>-0.226</td>
<td>416, 421</td>
</tr>
</tbody>
</table>
“By definition, complex networks are networks with more complex architectures than classical random graphs with their ‘simple’ Poissonian distributions of connections. The great majority of real-world networks... are complex ones. The complex organization of these nets typically implies a skewed distribution of connections with many hubs, strong inhomogeneity, and high clustering, as well as nontrivial temporal evolution. These architectures are quite compact..., infinitely dimensional small worlds.”
For decades, we tacitly assumed that the components of such complex systems as the cell, the society, or the Internet are randomly wired together. In the past decade, an avalanche of research has shown that many real networks, independent of their age, function, and scope, converge to similar architectures, a universality that allowed researchers from different disciplines to embrace network theory as a common paradigm. The decade-old discovery of scale-free networks was one of those events that had helped catalyze the emergence of network science, a new research field with its distinct set of challenges and accomplishments.

For decades, we tacitly assumed that the components of such complex systems as the cell, the society, or the Internet are randomly wired together. In the past decade, an avalanche of research has shown that many real networks, independent of their age, function, and scope, converge to similar architectures, a universality that allowed researchers from different disciplines to embrace network theory as a common paradigm. The decade-old discovery of scale-free networks was one of those events that had helped catalyze the emergence of network science, a new research field with its distinct set of challenges and accomplishments.

The decade-old discovery of scale-free networks was one of those events that had helped catalyze the emergence of network science, a new research field with its distinct set of challenges and accomplishments.
mainstream network science

- network = graph
- structure = connectivity
- function = statistical properties of ensemble
- architecture = graph topology
- robust to arbitrary rewiring that maintains statistical connectivity properties
- “disorganized” (Weaver)
“problems of simplicity”  
(Weaver 1948)

example: billiard balls

- classical dynamics provide exact descriptions of a small number of balls interacting on a table

“disorganized complexity” (Weaver 1948)

• “The physical scientists, with the mathematicians often in the vanguard, developed powerful techniques of probability theory and of statistical mechanics to deal with what may be called problems of disorganized complexity.”

• “The methods of statistical mechanics are valid only when the balls are distributed, in their positions and motions, in a helter-skelter, that is to say a disorganized, way.”
“organized complexity” (Weaver 1948)

- "For example, the statistical methods would not apply if someone were to arrange the balls in a row parallel to one side rail of the table, and then start them all moving in precisely parallel paths perpendicular to the row in which they stand. Then the balls would never collide with each other nor with two of the rails, and one would not have a situation of disorganized complexity."

Systems exhibiting organized complexity:
- biological systems (Weaver)
- ecosystems
- economies
- social systems
- advanced technologies (e.g., the Internet)
highly organized systems

• components are arranged in a very specialized way that enables functionality and/or robustness features
• even minimal random rearrangement of that structure tends to destroy its most salient features

claim:
• this structure is a consequence of specific constraints that are placed on functionality and/or behavior
• largely independent of the process by which this organization arises, whether by design or evolution.
Organization arises from constraints that are placed on functionality and/or behavior.
a constraint-based view of organization

“Hard Limits”

“design space”

Protocol-Based Architecture

Component constraints

System-level constraints

Fundamental claim: complex networks (that we care about) are the result of design (either evolution or engineering)
a constraint-based view of organization

- **System-level constraints**
  - Constraints on the system as a whole (e.g., functional requirements)

- **Component constraints**
  - Constraints on individual components (e.g., physical, energy, information)
• **Hard limits** on system characteristics
  • *implied by the intersection of component and system constraints*

• Most interesting when they do not follow trivially from the other constraints

• Examples:
  – Entropy/2\textsuperscript{nd} law in thermodynamics
  – Channel capacity theorems in information theory
  – Bode integral and related limits in control theory
  – Undecidability, NP-hardness, etc in computational complexity theory
• Emphasis on *protocols* (persistent rules of interaction) over *modules* (that obey protocols and can change)

• In reverse engineering,
  • figure out what rules are being followed
  • and how they govern system features or behavior

• In forward engineering,
  • specify protocols that ensure such system behavior
a constraint-based view of organization

“Hard Limits”

“design space”

Component constraints

System-level constraints

Protocol-Based Architecture
Arbitrary rewiring of highly organized systems destroys their most salient features.
mainstream network science

- network = graph
- structure = connectivity
- function = statistical properties of ensemble
- architecture = graph topology
- robust to arbitrary rewiring that maintains statistical connectivity properties
- “disorganized” (Weaver)

- formation: random growth, self-organization

organized complexity

- network = graph + data + ...
- structure = organization
- function = domain-specific performance, robustness
- architecture = protocols, constraints that deconstrain
- fragile to arbitrary rewiring (modularity = robust to structured rearrangement)
- highly organized

- formation: evolution, design, architecture
Network Centric Critical Infrastructures

Transportation

Essential Utilities

Telecommunications

Banking & Finance
Primary Activities:
1. *How infrastructure systems will respond* to major disruptions
   - *deliberate* (e.g., sabotage, vandalism, terrorism, war)
   - *non-deliberate* (accident, failure, natural disaster)

2. *how to invest limited resources* (for hardening, redundancy, or capacity expansion) to make these systems **resilient** to disruptions

Over 150 “Red Team” Case Studies

- **Energy**: Electric Power, Natural Gas, Petroleum Reserves
- **Transportation**: Roads & Bridges, Mass Transit, Ports
- **Data and Voice Communications**
- **Emergency Preparedness & Response**
- **Supply Chains**
- **Site Security**: Airports, Military Bases, Heads of State, Superbowl
- **Critical Project Management**

By viewing our critical infrastructure through the eyes of intelligent adversaries, we discover how systems can be extremely fragile, and how we can mount effective hardening and defensive efforts.
Using attack-based strategies to identify critical infrastructure components is not a new idea

“Scale-free” networks and the “Achilles’ heel” of the Internet


**CNN.com: Scientists spot Achilles heel of the Internet**

- An estimated three percent of nodes are down at an given time but no one notices because the system copes with it.

- "The reason this is so is because there are a couple of very big nodes and all messages are going through them. **But if someone maliciously takes down the biggest nodes you can harm the system in incredible ways. You can very easily destroy the function of the Internet,**" he added.

- "That's exactly the situation on the Internet: there are a couple of hubs that are crucial to the system," Barabasi explained.

Many networks with power laws don’t have an Achilles’ heel. The real Internet is only one such example.

No, many networks with power laws don’t have an Achilles’ heel. The real Internet is only one such example.

1. Graph connectivity only
2. Borrowed data, at face value
3. Inconsistent with engineering reality

The Internet is vulnerable to attacks on hubs.
Cascade-based attack vulnerability on the US power grid

Jian-Wei Wang, and Li-Li Rong

Institute of System Engineering, Dalian University of Technology, 2 Ling Gong Rd., Dalian 116024, Liaoning, PR China

Abstract

The vulnerability of real-life networks subject to intentional attacks has been one of the outstanding challenges in the study of the network safety. Applying the real data of the US power grid, we compare the effects of two different attacks for the network robustness against cascading failures, i.e., removal by either the descending or ascending orders of the loads. Adopting the initial load of a node \( j \) to be \( L_j = [k_j(\sum_{m} \Gamma_{jm})]^\alpha \) with \( k_j \) and \( \Gamma_{jm} \) being the degree of the node \( j \) and the set of its neighboring nodes, respectively, where \( \alpha \) is a tunable parameter and governs the strength of the initial load of a node, we investigate the response of the US power grid under two attacks during the cascading propagation. In the case of \( \alpha < 0.7 \), our investigation by the numerical simulations leads to a counterintuitive finding on the US power grid that the attack on the nodes with the lowest loads is more harmful than the attack on the ones with the highest loads. In addition, the almost same effect of two attacks in the case of \( \alpha = 0.7 \) may be useful in furthering studies on the control and defense of cascading failures in the US power grid.

Keywords: Cascading failure; Attack; US power grid; Critical threshold; Tunable parameter


Fig. 1. The scheme illustrates the correlation between the initial load of a node $i$ and its neighboring nodes, i.e., nodes $i_1,i_2,i_3$, and $i_4$.

$$L_j = \left[ k_j \left( \sum_{m \in \Gamma_j} k_m \right) \right]^\alpha$$

$k_j = \text{deg of node } j$

$\Gamma_j = \text{neighbors of node } j$

$\alpha = \text{tunable parameter}$

Fig. 2. The scheme illustrates the load redistribution triggered by an node-based attack. Node $i$ is removed and the load on it is redistributed to the neighboring nodes connecting to node $i$. Among these neighboring nodes, the one with the higher load will receive the higher shared load from the broken node.

$$\Delta L_{ji} = L_i \frac{L_j}{\sum_{m \in \Gamma_j} L_m}$$

1. graph connectivity only
2. borrowed data, at face value
3. inconsistent with engineering reality
main points

• networks ≠ graphs

• essential features: *connectivity* and *organization*
  – as key feature in complex engineering systems
  – and biological systems!
  – and social systems?

• architecture as “constraints that deconstrain”

• taking graph data at face value is dangerous
Scientists have always relied on hypothesis and experimentation. Now, in the era of massive data, there’s a better way.

“All models are wrong, and increasingly you can succeed without them.”
References


