Complex Structures and Collective Dynamics in Networked Systems

Foundations for Self-Adaptation and Self-Organization

Ingo Scholtes
Chair of Systems Design
ETH Zurich
Complex networks

“The network paradigm is shifting. From artifacts that we construct to phenomena that we study.”
(Christos Gkantsidis, Microsoft Research) [Gkantsidis et al., 2003]
Collective dynamics
Outline

I. Introduction to Network Science
II. Topological Properties of Complex Networks
III. Dynamical Processes in Networks
IV. Applications in Socio-Technical Systems
V. Temporal Networks
I. Introduction to Network Science

Image credit: Randall Munroe, xkcd.com
Network basics

A vertex or node

An edge or link

Undirected + unweighted

The adjacency matrix $A$ for a symmetric binary adjacency matrix is:

\[
\begin{bmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 
\end{bmatrix}
\]
Network basics

A vertex or node

An edge or link

directed + unweighted

A =

\[
\begin{bmatrix}
  a & b & c & d & e & f & g \\
  a & 0 & 1 & 0 & 0 & 0 & 0 \\
  b & 0 & 0 & 1 & 1 & 0 & 0 \\
  c & 1 & 0 & 0 & 0 & 0 & 0 \\
  d & 0 & 1 & 0 & 0 & 1 & 0 \\
  e & 0 & 0 & 0 & 0 & 0 & 1 \\
  f & 0 & 0 & 0 & 1 & 1 & 0 \\
  g & 0 & 0 & 0 & 1 & 0 & 1 \\
\end{bmatrix}
\]

asymmetric binary adjacency matrix
Network basics

- **vertex or node**
- **edge or link**

**A** = asymmetric (real) adjacency matrix

\[
A = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 3 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 3 & 1 & 0 & 2 \\
0 & 0 & 0 & 1 & 0 & 2 & 0
\end{bmatrix}
\]
Network basics

A vertex or node

An edge or link

A directed + unweighted graph

A matrix representation of the graph:

\[
A = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0
\end{bmatrix}
\]

- indegree = 2
- outdegree = 3
Network basics

**vertex or node**

**edge or link**

undirected + unweighted

\[ A = \begin{bmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 \\
\end{bmatrix} \]

indegree = 3
outdegree = 3
Network basics

A =

\[
\begin{pmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 \\
\end{pmatrix}
\]

vertex or node

edge or link

self-loop

indegree = 4
outdegree = 4

undirected + unweighted
Paths and diameter

\[ p(v, w) = (p_0, \ldots, p_l), p_i \in V, (p_i, p_i+1)E, p_0 = v, p_l = w \]

\[ \text{dist}(v, w) = \min_{p(v,w)} l(p(v, w)) \]
\[ \text{diam}(G) = \max_{v, w \in V} \text{dist}(v, w) \]
Paths and diameter

Elements of $A^n$ count paths of length $n$
Paths and diameter

A network is connected.

Diameter is three.

Network has two communities.

Algebraic methods capture elementary collective properties.
What can we say about this network?

Consistent Global States of Distributed Systems: Fundamental Concepts and Mechanisms

Özalp Babaoğlu\textsuperscript{1}  
Keith Marzullo\textsuperscript{2}

Technical Report UBLCS-93-1
January 1993

Abstract

Many important problems in distributed computing admit solutions that contain a phase where some global property needs to be detected. This subproblem can be seen as an instance of the Global Predicate Evaluation (GPE) problem where the objective is to establish the truth of a Boolean expression whose variables may refer to the global system state. Given the uncertainties in asynchronous distributed systems that arise from communication delays and relative speeds of computations, the formulation and solution of GPE reveal most of the subtleties in global reasoning with imperfect information. In this paper, we use GPE as a canonical problem in order to survey we illustrate the utility of the developed techniques by examining distributed deadlock detection and distributed debugging as two instances of GPE.
What can we say now?
And now?
And now?

8 links on average
Micro perspective

- Knowledge (control) of network details
  - Allows graph-theoretic analysis
  - Applied in protocol design, topology management, etc.
Macro perspective

- Knowledge (control) of aggregate properties
  - Uncertainty at microscopic level
  - Statistical regularities of all feasible topologies?

21 nodes
8 links
Network science = macro perspective

- **Modeling** of networked systems
  - We often only know aggregate statistics
  - Systems are highly dynamic
  - Assume randomness at the micro level
- **Design of probabilistic protocols**
  - Randomness is (often) cheap to implement
  - Utilize statistical regularities
  - Self-organization processes
- **Design of adaptive systems**
  - Systems can switch details on its own
  - System invariants

![Diagram with 7 nodes and 8 links]

Network science = macro perspective

22
What is a network model?

- Network model defines a class of networks
  - Network topology is an instance of this class
  - Model constrains possible instances
What is a network model?

- Network model = class of networks
  - Network topology = instance of this class
  - Model constrains possible instances
- Instance 1
What is a network model?

- Network model = class of networks
  - Network topology = instance of this class
  - Model constrains possible instances
- Instance 2
What is a network model?

- Network model = class of networks
  - Network topology = instance of this class
  - Model constrains possible instances
- Instance 3
Ingredients of a network model

- Model defines a macrostate
  - Example: degree sequence \{2,3,2,4,2,3,2\}

- Instance (graph) is a microstate \(r\)
  - \(r_1 = (1,2), (1,3), (2,3), (2,4), (4,5), (4,6), (4,7), (5,6), (6,7)\)
  - \(r_2 = (1,4), (1,6), (2,3), (2,4), (2,5), (3,7), (4,5), (4,6), (6,7)\)

- Network model defines probability space

\[
P(r) = \frac{w_r}{Z} \\
Z = \sum_r w_r
\]
Erdös-Rényi Graphs: $G(n, m)$ model

- $G(n, m)$ model
  - $n$ number of nodes
  - $m$ number of links

- Statistical weights

$$P(r) = \frac{1}{Z}$$

$$Z = \sum_r 1 = \binom{n}{2}$$

Erdös-Rényi Graphs: $G(n, p)$ model

- $G(n, p)$ model
  - $n$ number of nodes
  - $p$ probability for creating edge $(v, w)$

- Statistical weights
  - Probability of microstate $r$ depends on number of edges $m_r$

$$P(r) = p^{m_r} \cdot (1 - p)^{\binom{n}{2} - m_r}$$

Microstates of the $G(n, p)$ model

$n = 32$
$p = 0.04$

Collective properties of microstates?
Degree distribution of random graphs

- Microstate in $G(n, p)$ model defines degree sequence
- Degree of randomly chosen node in microstate $r$

\[ P(X = k) = p(k) = \binom{n-1}{k} p^k (1 - p)^{n-1-k} \]

- Moments of this (binomial) distribution

\[ \langle k \rangle = \sum_{k=0}^{\infty} k \cdot p(k) \quad \Rightarrow np \]
\[ \langle k^2 \rangle = \sum_{k=0}^{\infty} k^2 \cdot p(k) \quad \Rightarrow np(1 - p) \]
We deal with LARGE systems ... 

\[ p(k) = \binom{n-1}{k} p^k (1 - p)^{n-1-k} \]

\[ p(k) = \frac{\lambda^k e^{-\lambda}}{k!} \]

\[ n \rightarrow \infty \]

\[ \lambda := \langle k \rangle = np \equiv \text{const} \]

\[ p \rightarrow 0 \]

Poisson

\[ n \rightarrow \infty \]

Moivre-Laplace

\[ p(k) = N(\langle k \rangle, \langle k \rangle (1 - p)) \]

Poissonian approximation of degree distribution valid for sparse networks!
Connected components of random graphs
The birth of the giant component

$p_c = 0.01$
The birth of the giant component

\[ p_c = 0.005 \]

\[ np_c = 1 \]
Threshold phenomena in random structures

Images: Wikimedia Commons

Statistical mechanics is a branch of physics that provides abstractions and mathematical tools which allow to derive statements about the macroscopic bulk properties of materials based on incomplete information about the detailed behavior of atoms or molecules.

„materials“ ➔ read „networks“
„atoms or molecules“ ➔ read „computers, humans, animals, cells,“
Particle and network ensembles

- network \iff particle system
- node pair \iff volume element
- edge \iff particle
Statistical mechanics ...

Micro-canonical ensemble

\[ P_r(N, V, E) = \frac{1}{\Omega(N, V, E)} \]

All microstates \( r \) equiprobable

Grand-canonical ensemble

\[ P_r(\mu, V, T) = \frac{e^{-\frac{E_r - \mu N_r}{T} \frac{E_r - \mu N_r}{T}}}{Z(N, V, T)} \]

\[ Z(\mu, V, T) = \sum_r e^{-\frac{E_r - \mu N_r}{T}} \]

\( P_r \) depends on \( E_r \) and \( N_r \)
Statistical mechanics of complex networks

G(n,m) model

\[ P_r = \frac{1}{\Omega(n,m)} \]

\[ \Omega(n,m) = \binom{n}{2}^m \]

All microstates \( r \) equiprobable

G(n,p) model

\[ P_r = p^{m_r} \cdot (1 - p)^{\binom{n}{2} - m_r} \]

\( P_r \) depends on \( m_r \)

Statistical mechanics of complex networks

\[ P_G(n, p) = p^m \cdot (1 - p)^{\binom{n}{2} - m} \]

that is given by the \( G(n, p) \) model, in terms of the probability \( P_G(n, \mu, T) \) of the grand canonical ensemble. As a first step, we can substitute the thermodynamic quantities in equation 2.23 by their respective network analogies.

\[ P_G(n, \mu, T) := \frac{1}{Z(n, \mu, T)} \cdot e^{-\frac{E_G - \mu \cdot m_G}{T}} = \frac{1}{Z(n, \mu, T)} \cdot e^{\frac{m_G(\mu - \alpha)}{T}} \]

Based on equation 2.24, for the normalizing partition function \( Z(n, \mu, T) \) we then obtain:

\[ Z(n, \mu, T) := \sum_{G' \in G(n,p)} e^{-\frac{E_{G'} - \mu \cdot m_{G'}}{T}} = \sum_{G' \in G(n,p)} e^{\frac{m_{G'}(\mu - \alpha)}{T}} \]

In the latter equation, the summation goes over all possible networks with \( n \) nodes. Alternatively, we may consider the partition function as summing over all possible edge numbers. The maximum number of edges in a loop-free, labeled \( n \)-node network is \( \binom{n}{2} \). When changing the summation from running over all networks to running over all possible edge numbers, for each number of edges \( m \) we can further multiply the number of networks with \( m \) edges.
Reconsidering the $G(n, p)$ model

$$p(v, w) = \frac{1}{1 + e^{(e(v, w) - \mu)/T}}$$

- $T \to \infty$
  $$\Rightarrow p \to \frac{1}{2}$$
  all microstates equally likely

- $T \to 0$
  $$e(v, w) > \mu \Rightarrow p \to 0$$
  $$\mu > e(v, w) \Rightarrow p \to 1$$
  only one microstate possible
Equilibrium network dynamics

\[ p(v, w) = \frac{1}{1 + e^{\frac{e(v, w) - \mu(v, w)}{T}}} \]
Entropy of network ensembles …

- For macrostate $m$ defining a set of microstates $\Omega$

  \[
  H(m) := -\sum_{r \in \Omega} P_r \log P_R
  \]

- Captures our (lack of) information about network realizations

  $\Omega = \{r\}: P_r = 1 \Rightarrow H(m) = 0$

  $\forall r: P_r = \frac{1}{|\Omega|} \Rightarrow H(m) = \max$
Communication topologies

Highly structured → Statistical regularities → Self-Organization → Unstructured

- The 2nd law
- High entropy
- Self-Organization
- Statistical regularities
- Highly structured
- Unstructured

Performance of distributed algorithms
- High to Low

Maintenance overhead
- High to Low

Entropy
- Low to High
The physics world

collective properties

property

\[ P_x(A) \to 1 \]

statistical ensemble

macrostate

microdynamics

kinetic models

magnetism, state of matter

\[ V, N, \mu, T, E \]
Network science

collective properties

statistical ensemble

microdynamics

connected components diameter
average path lengths
search and routing efficiency
diffusion processes
resilience

|V|, |E|, P(k)

property

connection mechanism