Duality between static and dynamic networks

Dmitri Krioukov
Massimo Ostilli
CAIDA/UCSD

NIST-Bell Labs Workshop on Large-Scale Networks
October 2013
Motivation

• Real networks are growing, but static network models exist and are widely used
  – Simpler
  – More tractable

• In a recent work, a minimalistic growing extension of a “very static” model describes remarkably well the growth dynamics of some real networks (*F. Papadopoulos et al., Nature, 489:537, 2012*)

• Informally:
  – One camp: there cannot be any connection
  – Another camp: there should be some connection
    • Preferential attachment versus configuration model: what is the difference???
Graph ensemble

• Set of graphs $\mathcal{G}$
• With probability measure $P(G)$,

$$\sum_{G \in \mathcal{G}} P(G) = 1$$
Random graphs with hidden variables

- Given $N$ nodes
- Assign to each node a random variable $\kappa$ drawn from distribution $\rho_N(\kappa)$
- Connect each node pair $(s, t)$ with probability $p_N(\kappa_s, \kappa_t)$
Example: Classical random graphs

- Given $N$ nodes
- Hidden variables: None
- Connection probability: $p_N(\kappa_s, \kappa_t) = p$
- Results:
  - Degree distribution: Poisson ($\lambda = pN$)
  - Clustering: Weak
Example: Soft configuration model

• Given $N$ nodes

• Hidden variables:
  Expected degrees $\kappa$
  $\rho(\kappa) \sim \kappa^{-\gamma}$

• Connection probability:
  $p_N(\kappa_s, \kappa_t) = \frac{\kappa_s \kappa_t}{\kappa N}$

• Results:
  – Degree distribution:
    Power law ($P(k) \sim k^{-\gamma}$)
  – Clustering:
    Weak
Example: Random geometric graphs

• Given $N$ nodes
• Hidden variables:
  Node coordinates $\vec{\kappa}$ in a space $\rho_N(\vec{\kappa})$: uniform in the space
• Connection probability:
  $p_N(\vec{\kappa}_s, \vec{\kappa}_t) = \Theta(R - d_{st})$
• Results:
  – Clustering:
    Strong
  – Degree distribution:
    • Euclidean or spherical space: Poisson
    • Hyperbolic or de Sitter spaces: Power law

\[ d(\vec{\kappa}_s, \vec{\kappa}_t) < R \]
Probability measure

- Let
  - \( G = \{ a_{st} \} \) the adjacency matrix
  - \( \hat{\kappa} = \{ \kappa_1, \kappa_2, \ldots, \kappa_N \} \) a hidden variable assignment
  - \( p_{st} = p(\kappa_s, \kappa_t) \) the connection probability matrix

- Then
  - \( P(G|\hat{\kappa}) = \prod_{s<t} p_{st}^{a_{st}} (1 - p_{st})^{1-a_{st}} \)
  - \( \rho_N(\hat{\kappa}) = \prod_{t=1}^N \rho_N(\kappa_t) \)
  - \( P(G) = \int P(G|\hat{\kappa}) \rho_N(\hat{\kappa}) \, d\hat{\kappa} \)
Static versus dynamic graphs with hidden variables

- **Static ensemble $G_S$**
  - Given $N$ nodes
  - Assign to each node $t = 1, 2, \ldots, N$ a random variable $\kappa_t$ drawn from distribution $\rho_N(\kappa_t)$
  - Connect each node pair $(s, t)$ with probability $p_N(\kappa_s, \kappa_t)$

- **Dynamic ensemble $G_D$**
  - For each new node $t = 1, 2, \ldots$
  - Assign to $t$ a random variable $\kappa_t$ drawn from distribution $\rho_t(\kappa_t)$
  - Connect $t$ to each existing node $s$ with probability $p_t(\kappa_s, \kappa_t)$
Equivalence

• Let dynamic $t = 1, 2, ..., N$
• Two ensembles $\mathcal{G}_S$ and $\mathcal{G}_D$ are identical ($\mathcal{G}_S = \mathcal{G}_D$) if $P_S(G) = P_D(G)$ for all $G \in \mathcal{G}$
• Distributions $P_S(G) = P_D(G)$ if
  - $\rho_S(\tilde{\kappa}) = \rho_D(\tilde{\kappa})$ and
  - $p_S(\kappa_S, \kappa_t) = p_D(\kappa_S, \kappa_t)$
• The difference is
  - $\rho_N(\kappa_t)$ (static) versus $\rho_t(\kappa_t)$ (dynamic)
  - $p_N(\kappa_S, \kappa_t)$ (static) versus $p_t(\kappa_S, \kappa_t)$ (dynamic)
Weak equivalence

• Ensembles $\mathcal{G}_S$ and $\mathcal{G}_D$ are weakly equivalent if $\mathcal{G}_S = \mathcal{G}_D$ for some $N$

• The simplest example
  – Dynamic $\rho_t(\kappa_t)$ is equal to static $\rho_N(\kappa_t)$
  – Dynamic $p_t(\kappa_s, \kappa_t)$ is equal to static $p_N(\kappa_s, \kappa_t)$
  – The difference is only in node labeling
Strong equivalence

- Ensembles $G_S$ and $G_D$ are strongly equivalent if $G_S = G_D$ for any $N$
- In this case the connection probability cannot depend on graph size $N$
  - $p_N(\kappa_s, \kappa_t) = p_t(\kappa_s, \kappa_t) = p(\kappa_s, \kappa_t)$
- The simplest example
  - $\rho_N(\kappa_t) = \rho_t(\kappa_t) = \rho(\kappa_t)$
  - Graphs are dense because average degree
    \[ \bar{k} = N \iint \rho(\kappa) p(\kappa, \kappa') \rho(\kappa') \, d\kappa \, d\kappa' = O(N) \]
- In sparse strongly equivalent ensembles, the connection probability does not depend on graph size, but the hidden variable distribution does
Example: Classical random graphs

• Ensemble $\mathcal{G}_{N,p}$ is strongly equivalent and dense since \( \bar{k} = Np \)
  - The growing definition is: connect new nodes \( t \) to existing nodes \( s < t \) with probability \( p \)

• An attempt to fix: ensemble $\mathcal{G}_{N,\bar{k}}$:
  - Static: connect all node pairs with probability \( p = \frac{\bar{k}}{N} \)
    - The ensemble is $\mathcal{G}_{N,p}$
    - The degree distribution is Poisson
  - Dynamic: connect new nodes \( t \) to existing nodes \( s < t \) with probability \( p = \frac{\bar{k}}{t} \)
    - The ensemble is not $\mathcal{G}_{N,p}$
    - The degree distribution is exponential

• The ensembles are very different because the connection probability depends on graph size
Example: Soft configuration model

- Strong equivalence in the dense case (graphons)
- No strong equivalence in the sparse case since the connection probability $p_{st} \sim \frac{\kappa_s \kappa_t}{N}$ depends on graph size
Example: Preferential attachment

• Weak equivalence in the soft formulation (M. Boguna et al., PRE 68:36112, 2003)

• No strong equivalence since the probability that new node $t$ connects to existing node $s$
  $$p_{st} \sim \frac{k_s}{\sum_s k_s}$$
  depends on graph size

• Needless to say, $P(G)$ is unknown
Example: Random geometric graphs

- **Strong equivalence in sparse graphs!**
- Connection probability $p_{st} = \Theta(R - d_{st})$ does not depend on graph size
- The distribution of new node coordinates $\rho_t(\kappa_t)$ is such that $\rho_S(\vec{\kappa}) = \rho_D(\vec{\kappa})$ because both static and dynamic constructions implement the same (Poisson) point process
Poisson point processes behind random geometric graphs

- Let $r_t$ be the radial coordinate of node $t$
- Then $V_t = |B(0, r_t)|$ is its volume coordinate
- Points $\{V_t, t = 1,2, \ldots \}$ is a PPP on $\mathbb{R}_+$
Two pairs of identical ensembles

• Fixed-volume pair
  – Static ensemble $G_{S,V}$
    • Given $V$, sample $N$ from the Poisson distribution with mean $\delta V$, where $\delta$ is the point density (PPP rate)
    • Sample $N$ points on $[0, V]$
  – Dynamic ensemble $G_{D,V}$
    • Given $\Delta V$, sample $N$ from the Poisson distribution with mean $\delta \Delta V$, where $\delta$ is the point density (PPP rate)
    • Sample $N$ points on $[V, V + \Delta V]$

• Fixed-size pair
  – Static ensemble $G_{S,N}$
    • Given $N$, sample $V$ from the Gamma distribution with rate $\delta$ and shape $N$
    • Sample $N - 1$ points on $[0, V]$
  – Dynamic ensemble $G_{D,N}$
    • Given $\Delta N$, sample $\Delta V$ from the Gamma distribution with rate $\delta$ and shape $\Delta N$
    • Sample $\Delta N - 1$ points on $[V, V + \Delta V]$
Two pairs of identical ensembles

\[ G_{D,N} \quad \cdots \quad V \quad N \quad N+1 \quad V+V' \quad N+N' \]

\[ G_{D,V} \quad \cdots \quad V \quad N+1 \quad V+V' \quad N+N' \]

\[ \mathcal{G}_{S,V} = \mathcal{G}_{D,V} \]

\[ \mathcal{G}_{S,N} = \mathcal{G}_{D,N} \]
Final example: de Sitter causal sets

- Random geometric graphs in de Sitter spacetime
  - Angular coordinates $\theta$ are as in all spherically symmetric random geometric graphs, i.e., sampled from the uniform distribution on a unit sphere $\mathbb{S}^d$
  - “Radial” (conformal time) coordinates $\eta$ are sampled from $\rho(\eta) \sim \sec^{d+1} \eta$
  - Nodes are connected if $\Delta \eta > \Delta \theta$

- The resulting graph ensembles are
  - Strongly equivalent
  - Sparse
  - Have power-law distributions of node degrees with exponent $\gamma = 2$
  - Have strong clustering
The graphs show the distribution of node degrees in different networks.

**Graph (a):**
- **P(k)**: Probability distribution of node degree k.
- **Node Degree k**: The x-axis represents the node degree k.
- **Graph Elements**:
  - Internet
  - Brain
  - Trust
  - De Sitter dyn.
  - De Sitter stat.

**Graph (b):**
- **Clustering C(k)**: Clustering coefficient distribution of node degree k.
- **Node Degree k**: The x-axis represents the node degree k.
- **Graph Elements**:
  - Internet
  - Brain
  - Trust
  - De Sitter dyn.
  - De Sitter stat.
Conclusions

• There does exist a very simple connection between equilibrium and non-equilibrium approaches to network analysis

• Some real networks appear to be “close” to graph ensembles that admit dual, equilibrium and growing, description

• This justifies applications of powerful equilibrium tools to the analysis of real networks

• Does nature optimize?

• If the least action principle applies, then what is the action?
• D. Krioukov and M. Ostilli
  Duality between equilibrium and growing networks,
  *PRE*, v.88, p.022808, 2013