Social Distance and Network Structures

Ryota Iijima (Univ. of Tokyo)
Yuichiro Kamada (Harvard Univ.)
1 Introduction

- Different networks have different structures, which are characterized by: Clustering coefficient ($Cl$), average path length ($APL$), and degree distribution etc.
  - In large $Cl$ networks, one’s neighbors tend to be neighbors themselves.
  - Small $APL$ networks connect agents by short paths.

- **E-mail network** ($n = 59,814$; Ebel et al., 2002):
  $Cl = .0344, APL = 4.95$;

- **Web hyperlink network** ($n = 153,127$; Adamic, 1999):
  $Cl = .1078, APL = 3.1$;

- **Coauthorship network in economics** ($n = 81,217$; Goyal et al., 2006):
  $Cl = .16, APL = 9.5$.

- Why different networks have different structures?
1.1 Motivation of the Paper

- One of the important aspects in the formation of networks.
  - Each agent is heterogeneous with respect to his “type”.

- Literature that models heterogeneous agents’ types.
    - Agents are partitioned into several groups or categories.

- Johnson and Gilles (2000)
  - Network formations based on geographical cost patterns. (uni-dimensional type space)
1.2 Motivation of the Paper –cont’d.

- The existing literature is not suitable for situations where
  - agents have multiple aspects of characteristics
  - agents belong to multiple groups.

- We employ *multi-dimensional type space*.

  → But how do the agents integrate and evaluate the information about the relationships?

- We suppose that in different networks, the agents have *different ways of measuring distances* between them.
1.3 Networks

- $N$ is a set of agents. $\#N = n$.

- $g$ is a network: $ij \in g$ iff there exists a link between agents $i$ and $j$.

- $q_i(g)$ is agent $i$'s degree, i.e. the number of $i$'s neighbors. $q_i(g) = \#\{j | ij \in g\}$.

- $PL_{ij}(g)$ is path length between $i$ and $j$, i.e. the smallest number of links needed to connect $i$ and $j$.

  $APL(g)$ is average path length, the average of $PL_{ij}(g)$'s over all pair $ij$'s that are connected in $g$.

- $Cl_i(g)$ is agent $i$'s clustering, the probability that two neighbors of agent $i$ are also linked.

  $Cl(g)$ is clustering coefficient, is the average of $Cl_i(g)$ of all the $i$'s.
1.4 Example: 15’th-century Florentine Marriages

\[ PL_{\text{Peruzzi,Medici}} = 3 \]

\[ Cl_{\text{Medici}} = \frac{1}{15} \]

From Padgett and Ansell (1993).
2 Definitions

2.1 Type Space and Social Distances

- Agents use social distance when they evaluate the value of relationship from others.

- Each agent is located on a point in a type space, normalized to $X = [0, 1]^m$. So, agent $i$'s type is an $m$-dimensional vector: $x_i = (x_{i1}, ..., x_{im}) \in X$. Each dimension describes his affiliation, location, taste, etc.

- Assume that $\{x_1, x_2, \cdots, x_n\}$ follow i.i.d and continuous dist. $f$ over $X$. Each dimension does not correlate with each other.

- We consider various ways to measure distances in the type space. (These don't necessarily satisfy usual axioms of metrics.)

- First, consider two special cases:
  - *Max norm*: $d_{\text{max}}(i, j) = \max_{1 \leq h \leq m} \{|x_{ih} - x_{jh}|\}$,
  - *Min norm*: $d_{\text{min}}(i, j) = \min_{1 \leq h \leq m} \{|x_{ih} - x_{jh}|\}$. 
More generally, consider a class of social distance, $k^{th}$ norm:
\[ d^{(k)}(i, j) = |x_{il} - x_{jl}| \text{ s.t. } \# \{h : |x_{ih} - x_{jh}| < |x_{il} - x_{jl}| \} = k - 1 \]

In words, $k^{th}$ smallest dimension-wise distance. (Assuming that there is no tie in these distances).

Example: $m = 4$ and two agents $i$ and $j$, $x_i = (0.3, 0.2, 0.4, 0.6)$ and $x_j = (0.7, 0.7, 0.7, 0.7)$.
$\rightarrow$ Dimension-wise distances are $(0.4, 0.5, 0.3, 0.1)$.

- If we use 1st norm, then $d^{(1)}(i, j) = 0.1$;
- If we use 2nd norm, then $d^{(2)}(i, j) = 0.3$;
- If we use 3rd norm, then $d^{(3)}(i, j) = 0.4$;
- If we use 4th norm, then $d^{(4)}(i, j) = 0.5$.

Rough interpretation: If $k$ is large, agents (have to) care a lot of aspects of others’ types; If $k$ is small, then they don’t (have to) care a lot of aspects of other’s types.
3 The Model and Terminology

- Agent’s payoff is composed of benefit and cost from his neighbors,
  \[ u_i(g) = \sum_{j \in N_i(g)} b(d^{(k)}(i, j)) - c(q_i), \]
  where

  \( b(\cdot) \): benefit is weakly decreasing and continuous from the left.
  \( c(\cdot) \): cost is strictly increasing.

- \( g \) is pairwise stable (Jackson and Wolinsky 1996) if no agent has an incentive to delete a link in \( g \), and no unlinked pair has an incentive to form a link.

**Proposition 1.** Assume \( c \) is linear. Then, a pairwise stable network \( g \) is unique, and is generated by **cutoff-rule**. That is, there exists \( \hat{d} \) such that
  \[ g = g(\hat{d}) := \{ i, j \mid d^{(k)}(i, j) \leq \hat{d} \} \]

- Link \( ij \) is surely formed iff the distance between them is not larger than the **cutoff value** \( \hat{d} \). We focus on this type of networks, \( g(\hat{d}) \), assuming linear \( c \).
4 Cutoff-Rule Model

4.1 Clustering Coefficient

- We focus on the clustering coefficient in the limit as $\hat{d}$ going to zero:
  \[
  Cl^* := \lim_{\hat{d} \to 0} \lim_{n \to \infty} E[Cl(g(\hat{d}))].
  \]

- We want to consider large networks, and to avoid difficulty in calculations.
- We want to compare the implication of different $k$’s. Denote $Cl^*(k, m)$ instead of $Cl^*$.

**Proposition 2.** For each $m$ and $k \leq m$,

\[
Cl^*(k, m) = \left( \frac{m}{k} \right)^{-1} \left( \frac{3}{4} \right)^k.
\]

**Corollary 1.** $Cl^*(k, m)$ is decreasing in $m$.

**Corollary 2.** $Cl^*(k, m)$ is decreasing in $k$ for small $k$, and increasing for large $k$.

**Corollary 3.** If $m < 9$, $Cl^*(m, m) > Cl^*(1, m)$ holds.
That is, $Cl^*$ is higher with Max norm than with Min norm.
4.2 Example: Max norm vs Min norm when $m = 2$

- What is the prob. that $jk \in g$ given $ij, ik \in g$?

- Max norm satisfies the triangle inequality.
  $\xrightarrow{\text{ }} d(j, k)$ is bounded above by $d(i, j) + d(i, k)$.

- But in the Min norm case, $d(j, k)$ can be much larger than $d(i, j), d(i, k)$.
4.3 Average Path Length

- We focus on the limit of the value: \( APL^* := \lim_{\hat{d} \to 0} \lim_{n \to \infty} E[APL(g(\hat{d}))] \).
  We also use \( APL^*(k, m) \) as before.

**Proposition 3.** Take any \( k \) and \( m \) such that \( k < m \). Then, \( APL^*(k, m) \) is
\[
\left\lfloor \frac{m}{m-k} \right\rfloor + 1 \text{ if } \frac{m}{m-k} \text{ is not an integer and } \frac{m}{m-k} \text{ if it is, where } \left\lfloor \cdot \right\rfloor \text{ is Gaussian.}
\]

**Corollary 4.** If \( k < m \), \( APL^*(k, m) \) is decreasing in \( m \), and is increasing in \( k \).

- \( APL \) is small if the type space is rich (if \( m \) is large), and/or if agents don’t care a lot of aspects of the others (if \( k \) is small).

- Proposition 3 rules out the case of Max norm, where \( k \) is exactly equal to \( m \).

**Proposition 4.** \( APL^*(m, m) = \infty \).

- Here, we see a striking difference: Max norm satisfies the triangle inequality. Then, \( APL \) goes to infinity, since it becomes hard to find short path between the agents.
4.4 Consistency with the Stylized Facts

- When $k < m$, our model has the properties well observed in social networks.
  - smaller $APL$ compared with lattice networks.
    $\leftrightarrow$ “small world” networks.
  - larger $Cl$ compared with random (Poisson) networks.
5 Conclusion

- We proposed a network formation model, which provides an explanation about why different networks have different structures.

- We considered multi-dimensional “types” of agents, and various ways of measuring distance in the type space.

- Under linear costs, it suffices to focus on the cutoff-rule to analyze pairwise stable networks.

- Clustering coefficient and average path length are characterized in terms of $k$ and $m$.

- For nonlinear $c$, pairwise stable network may not be neither unique nor generated by cutoff-rule. Show that “strongly stable” network is unique, and also generated by the cutoff-rule (w/ heterogeneous cutoff values).