Network bottleneck and speed of learning

Prepared for class presentation (SS211)

Jin F. Huang

May, 2013

Motivation

- How fast does information travel within a network?
- How long does it take for a community to reach consensus?
- How do we arrange a communication network so that it is more conducive to forming compromises?

Learning environment

- DeGroot's model on learning
- Linear updating, repeated learning from neighbors
- Updating matrix

$$T = 1/2(I + \Delta^{-1}A),$$

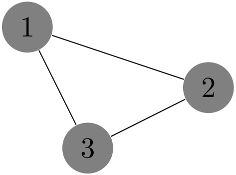
  where $\Delta = diag(d_1, \cdots, d_n)$, I is the identity matrix, and A is the adjacency matrix.

  In words, i gives his own opinion 1/2 weight and the rest evenly distributed among his neighbors.

- Updating rule

$$b_i^{t+1} = \sum_{j \in N} T_{ij} b_j^t$$

An example



$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \qquad \Delta = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \qquad T = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

Convergence of long run behavior

- T has a unique stationary distribution π,

$$\pi_i = \frac{d_i}{\sum_j d_j}.$$

- For any $b^0 \in [0,1]^N$,

$$b^\infty \triangleq T^\infty b^0 = \Big(\sum_i \pi_i b_i^0\Big)\mathbf{e},$$

where $\mathbf{e}$ is $(1, \cdots, 1)^{tr}$.

In words, long run behavior converges. The convergent behavior is a weighted average of the initial behavior.

Consensus time

Fix an initial behavior b_0, then $b_i^\infty = \sum_j \pi_i b_j^0$.

- Distance of two probability distributions (μ and ν)

$$||\mu - \nu|| \stackrel{\Delta}{=} \max_{S \subseteq N} |\mu(S) - \nu(S)|.$$

- Consensus distance

$$cd(t;T) \stackrel{\Delta}{=} \max_{i \in N} ||T_i^t - \pi||$$

- We look at $ct(t;T)$ because whenever $cd(t;T) < \epsilon$,

$$|b_i^t - b_i^\infty| = |\sum_j T_{ij}^t b_j^0 - \pi_j b_j^0| \leq \sum_j |T_{ij}^t - \pi_j| b_j^0 < \epsilon ||b^0||$$

Consensus time, cont.

Consensus time

$$ct(\epsilon; T) = \inf \{t \geq 0 : cd(t; T) < \epsilon\}$$

In words, this is the amount of time that the updated behavior b^t is ϵ-close to the long run, steady-state behavior b^∞.

Consensus time and spectral gap

- The spectral gap

$$\gamma \triangleq \lambda_1 - \lambda_2 = 1 - \lambda_2$$

Proposition

$$-log(2\epsilon)(\frac{1}{\gamma} - 1) \leq ct(\epsilon; T) \leq -log(\pi_{min}\epsilon)\frac{1}{\gamma}$$

Bottleneck ratio

- Network influence j has on i

$$q(i,j) \triangleq \pi_i T_{i,j}$$

- Influence that group S_1 has on S_2

$$q(S_1, S_2) = \sum_{(i,j)\in(S_1,S_2)} \pi_i T_{i,j}$$

- Bottleneck ratio of group S

$$\Phi(S) \triangleq \frac{q(S, S^C)}{\pi(S)}$$

Bottleneck ratio, cont.

- Bottleneck ratio of the network

$$\Phi \triangleq min\{\Phi(S) : S \subseteq N, \pi(S) \leq 1/2\}$$

- It measures how much the critical group S^* is isolated from the rest of the network

- Φ is between 0 and $\frac{1}{2}$

- $\Phi = 1/2$ for a triangle; $\Phi = 1/n$ for a circle of even size

Bottleneck ratio and spectral gap

Proposition

$\frac{\Phi^2}{2} \leq \gamma \leq 2\Phi$

Corollary

$-log(2\epsilon)(\frac{1}{2\Phi} - 1) \leq ct(\epsilon; T) \leq -log(\pi_{min}\epsilon)\frac{2}{\Phi^2}$

- A large bottleneck ratio is helpful for fast information propagation

A geometric view of Φ

- Let $e(S, S^c)$ be the number of cross edges between the group S and its complement

- Φ can also be expressed as

$$\Phi(S) = \frac{1}{2} \frac{e(S, S^c)}{\sum_{i \in S} d_i}$$

- A group would act as blockade to information exchanges if it is *large* and has very few cross links

Speed of learning for large networks

- $ct(\epsilon; T)$ is a measure of a fixed network

- What if we want to estimate the rate of learning of network that we don't its precise structure but we know its generating process

- We look at consensus time with respect to its size

$$ct(n) \triangleq min\{t \geq 0 : cd(t; T(n)) < 1/(2e)\},$$

where e is the natural number

Speed of learning for large networks, cont.

- $ct(n)$ is the amount of time it takes in a network of size n to reach a certain level of closeness within b^∞

Lemma

$$ct(\epsilon; T(n)) \leq log(\epsilon^{-1})ct(n)$$

Examples of large non-random networks

- A complete network, $\Phi = 1/4$. Then, $ct(n) \leq C \log n$
- A star, $\Phi = 1/2$. Then, $ct(n) \leq C \log n$
- A circle, $\Phi = 1/n$. Then, $ct(n) \leq Cn^2 \log n$
- A dumbbell, $\Phi = 1/n^2$. Then, $ct(n) \leq Cn^4 \log n$

Erdös Rényi

Erdös-Rényi $G(n, \lambda/n)$ mean that each one of the $\binom{n}{2}$ links would be deleted with probability $1-p$ independently. If $p > 1/n$, then there is a giant component.

Theorem

Let the network be the Erdös-Rényi $G(n, \lambda/n)$ with $\lambda > 1$. Then for a large enough n, consensus time in the largest component is almost surely

$$ct(n) \leq Clog^2(n).$$

Theorem

Let the network be the Erdös-Rényi $G(n, \lambda \log n/n)$ with $\lambda > 1$.

$$ct(n) \leq C \log n.$$

Preferential attachment

$PA(n, m)$: Each new node is added and connects to m neighbors. The choice of getting connected for each existing node is proportionate to its degree.

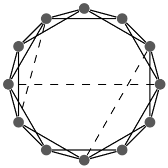
Theorem

Let the network be the Erdös-Rényi $G(n, \lambda \log n/n)$ *with* $\lambda > 1$.

$$ct(n) \leq C \log n.$$

Small world

$NW(n, k, p)$: Starting from a circle with each agent connecting to his closest $2k$ neighbors, a link is added to each non-linking pair with probability p.



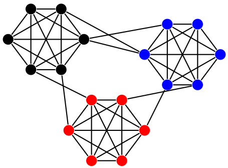
Theorem

Let the network be $NW(n, k, \lambda/n)$ with $\lambda > 1$. Then

$$ct(n) \leq Clog^2(n).$$

Island model

$IM(n, K, p_s, p_d)$: K types, p_s is the internal connecting probability, and p_d is between types



Theorem

Let $\frac{p_d}{p_s} = \frac{\lambda}{n^a}$. *If* $a = 0$ *and* $0 < \lambda < 1$, *then* $ct(n) \leq C \log n$. *If* $a, \lambda > 0 \geq 0$, *then* $ct(n) \leq Cn^{2a} \log n$.

Conclusion

- Φ is easy to interpret, visualize, and estimate
- Φ can be used to analyze large, random networks
- Limitations
  - ▷ updating is mechanical
  - ▷ the bounds are not tight (for example, a dumbbell); λ_2 is a more precise measure of speed
  - ▷ lower bound is mostly missing