Parsimonious Flooding in Dynamic Networks

Pierre Fraigniaud
CNRS and University Paris Diderot

Joint work with H. Baumann (U. Paris Diderot) and P. Crescenzi (U. Florence)
To appear in Proc. PODC 2009

Projet A.N.R. "ALADDIN"
La Rochelle, June 11-12, 2009
Dynamic graphs

Applications/motivations

- Peer-to-peer networks: Users join and leave at will
- Wireless mobile networks: Users move
- Internet, the WWW,…
- Social networks (human, animal,…)
- etc.

Main issue:

- Does dynamism help?
- If yes, in what extend? at which cost?
Ups and downs
Random graphs revisited

One possible dynamic graph definition:

$$G = (G_0, G_1, G_2, \ldots) \text{ where } G_t \in \mathcal{G}_{n,p} \text{ for all } t \geq 0$$

Sequences of random graphs

- capture abrupt modifications of the network structure
- but do not capture time dependencies
Evolving graphs

**Definition (Ferreira)**

An **evolving** graph is an infinite sequence $G = (G_0, G_1, G_2, \ldots)$ of graphs on the same vertex set $V$.

Evolving graphs are

- interesting for the design of algorithms (ex.: connectivity, flow, etc.) for dynamic networks: some polynomial problems become NP-hard;
- but
  - too general to enable the design of reasonably efficient distributed protocols,
  - and still do not capture time dependencies.
Markovian evolving graphs

Let $G$ be a set of graphs with the same vertex set $V$

Let $P$ be a probability transition matrix.

**Definition (Avin, Koucký and Lotker)**

A Markovian Evolving Graph is a Markov chain $(G, P)$:

$$\Pr[G_{t+1} = G_{t+1}^* | G_t = G_t^*, \ldots, G_0 = G_0^*] = \Pr[G_{t+1} = G_{t+1}^* | G_t = G_t^*]$$

$$\Pr[G_{t+1} = G_{t+1}^* | G_t = G_t^*] = P[G_{t+1}^*, G_t^*]$$

**Example:** Mobility model (Random Way Point)
Definition (Clementi, Pasquale, and Silvestri)

- Node velocity $\rho$: next position chosen u.a.r. in $B(\rho)$
- Transmission radius $r$: UDG model (can assume $r = 1$)
Edge Markovian evolving graphs

Definition (Clementi, Macci, Monti, Pasquale, and Silvestri)

Edge-Markovian process $\mathcal{M}_{n,p,q}$

- Birth-rate $p$ and death-rate $q$, $0 < p < 1$, $0 < q < 1$
- $\mathcal{M}_{n,p,q}$ generates $(G_0, G_1, G_2, \ldots)$ with $V(G_t) = \{1, 2, \ldots, n\}$
  - if $e \notin E(G_{t-1})$ then $e \in E(G_t)$ with probability $p$;
  - if $e \in E(G_{t-1})$ then $e \notin E(G_t)$ with probability $q$. 

![Graph example](image)
Stationary distribution

Recall:

- A Markov chain is **irreducible** if it is possible to get to any state from any state;
- A state $S$ is **recurrent** if the probability that the chain returns to $S$ is 1;
- A state $S$ is **positive recurrent** if the expected time before the chain returns to $S$ is finite.

An irreducible chain has a stationary distribution if and only if all of its states are positive recurrent (and then the stationary distribution is unique).

Stationary: $\mathcal{G}_{n,\hat{p}}$ with $\hat{p} = \frac{p}{p + q}$

Remark

- $G_0 \in \mathcal{G}_{n,\hat{p}} \Rightarrow G_t \in \mathcal{G}_{n,\hat{p}}$ for any $t \geq 0$.
- $\Pr[G_{t+1} \mid G_t]$ is specified by $p$ and $q$. 
Flooding

In **static** graphs:

- **step t**
- **step t+1**
- **step t+2**

In **dynamic** graphs:

- **step t**
- **step t+1**
- **step t+2**
Impact of dynamism

- Does dynamism help?
- If yes, in what extend? at which cost?

Previous work

- Clementi et al., 2008
  - $\forall G_0, p, q$, the flooding time in edge-Markovian graphs is at most $O(\log n \log (1 + np))$, w.h.p.
  - For $E(G_0) = \emptyset$, for any $0 < p, q < 1$, the flooding time is at least $\Omega(\log n / np)$, and if $p \geq c \log n / n$ for $c > 1$ then the flooding time is at least $\Omega(\log n / \log (1 + np))$.

- Clementi at al., 2009
  - If $G_0 \in G_{n, \hat{p}}$ and $\hat{p} \geq c \log n / n$ with $c$ large enough, the flooding time is at most $O\left(\frac{\log n}{\log np} + \log \log np\right)$ and at least $\Omega\left(\frac{\log n}{\log np}\right)$. 
**Flooding time [Baumann, Crescenzi, F.]**

Tight bounds on flooding time, \( \forall p, q \), for \( G_0 \in G_{n, \hat{p}} \)

<table>
<thead>
<tr>
<th>( 0 &lt; \hat{p} \leq \frac{c}{n}, c &gt; 0 )</th>
<th>( \frac{1}{n} \ll \hat{p} \leq \frac{c \log n}{n}, c &lt; 1 )</th>
<th>( \hat{p} \geq \frac{c \log n}{n}, c &gt; 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flooding time</td>
<td>( \Theta\left( \frac{\log n}{np} \right) )</td>
<td>( \Theta\left( \frac{\log n}{np} \right) )</td>
</tr>
</tbody>
</table>

\[ \text{np} \leq \log \hat{np} \quad \text{np} \geq \log \hat{np} \]

---

1Our results holds a.a.s. (not necessarily w.h.p.).
Parsimonious flooding

Definition
A $k$-active flooding protocol forwards a message during $k$ time steps.

Informed nodes at time $t + 1$:

$$I_{t+1}^{(k)} = I_t^{(k)} \cup N_t^{(k)}$$

where $N_t^{(k)}$ is the set of all nodes that are neighbors in $G_t$ of at least one node in $I_t^{(k)} \setminus I_{t-k}^{(k)}$.

$$T_s^{(k)} = \min\{t \geq 0 \mid I_t^{(k)} = [n]\}$$

Definition
The reachability threshold for the flooding protocol in $M_{n,p,q}$ is the smallest integer $k$ such that $T_s^{(k)} < \infty$ a.a.s., for any $s \in [n]$. 
### Main results [Baumann, Crescenzi, F., PODC ’09]

<table>
<thead>
<tr>
<th>(0 &lt; \hat{p} \leq \frac{c}{n}, c &gt; 0)</th>
<th>(\frac{1}{n} \ll \hat{p} \leq \frac{c \log n}{n}, c &lt; 1)</th>
<th>(\hat{p} \geq \frac{c \log n}{n}, c &gt; 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flooding time</td>
<td>(\Theta\left(\frac{\log n}{np}\right))</td>
<td>(\Theta\left(\frac{\log n}{np}\right))</td>
</tr>
<tr>
<td>Reachability</td>
<td>(\Theta\left(\frac{\log n}{np}\right))</td>
<td>(\Theta\left(\frac{\log n}{np}\right))</td>
</tr>
</tbody>
</table>

![Diagram showing the flooding time and reachability for different ranges of \(\hat{p}\).](image)

\(\hat{p} = p / (p+1)\)

\(c \log(1+n \hat{p}) = np\)
Auxiliary graphs

Let $a$ and $b$ be two real numbers such that $0 < a, b < 1$.

Let $Z$ r.v. defined by

$$\Pr[Z = i] = \begin{cases} a & \text{if } i = 1, \\ (1 - a)(1 - b)^{i-2}b & \text{if } i > 1. \end{cases}$$

**Definition (weighted random graphs $G_{n,a,b}$)**

Node set is $[n]$; Each of the $\binom{n}{2}$ edges $e$ is given a weight $Z_e$, independently from the other edges, where $Z_e \sim Z$.

**Definition ($k$-bounded weighted random graphs $G_{n,a,b}^{(k)}$)**

Node set is $[n]$; An edge $e$ is present only if $Z_e \leq k$, and whenever the edge $e$ is present it receives the weight $Z_e$. 
Reduction lemma

For any $s \in [n]$, let $X_s^{(k)} =$ eccentricity of $s$ in $G_{n,\hat{p},p}^{(k)}$.

For any $n \geq 1$, any $0 < p, q < 1$, any $k \in \mathbb{N}^+ \cup \{\infty\}$, and any $s \in [n]$, we have:

**Lemma**

$T_s^{(k)} \sim X_s^{(k)}$, that is $\Pr(T_s^{(k)} = x) = \Pr(X_s^{(k)} = x)$.

**Note:** Probability space for the l.h.s. of the equality is $\mathcal{M}_{n,p,q}$, while the one for the r.h.s. is $G_{n,\hat{p},p}^{(k)}$. 
The three regimes of random graphs $G_{n,p}$

- $p < \frac{1}{n}$: No giant component
- $p = \frac{\ln(n)}{n}$: A unique giant component
- $p > \frac{\ln(n)}{n}$: Connected
A lower bound

**Lemma**

If $p \to 0$ and $\hat{p} \to 0$ when $n \to \infty$, then the reachability threshold is, a.a.s., at least $\Omega\left(\frac{\log n - n\hat{p}}{np}\right)$.

**Proof.**

Let $\pi = \Pr[w(e) \leq k]$.

We have $\pi = 1 - (1 - \hat{p})(1 - p)^{k-1}$.

For flooding to complete a.a.s., we must have $\pi \geq \frac{\log n}{n}$.

Therefore,

$$k \geq 1 + \frac{\log(1 - \frac{\log n}{n}) - \log(1 - \hat{p})}{\log(1 - p)}$$

$$= 1 + \frac{\log n - n\hat{p}}{np}(1 + o(1)).$$
Beyond the connectivity threshold

\[ n\hat{p} - \log n \to \infty \]

**Theorem**

\[ T_s^{(1)} \leq O\left(\frac{\log n}{\log(n\hat{p})}\right) \text{ a.a.s.} \]

**Proof.**

Reduction Lemma: \( \Pr\{ T_s^{(1)} \leq x \} \leq \Pr\{ \text{diam}(G_{n,\hat{p},p}^{(1)}) \leq 2x \} \).

Now, \( G_{n,\hat{p},p}^{(1)} = G_{n,\hat{p}} \).

If \( n\hat{p} - \log n \to \infty \), then \( G_{n,\hat{p}} \) is a.a.s. connected [Bollobás].

In fact, a.a.s., \( \text{diam}(G_{n,\hat{p}}) \leq O\left(\frac{\log n}{\log(n\hat{p})}\right) \) [Chung and Lu, 2001]

Therefore, a.a.s., \( T_s^{(1)} \leq O\left(\frac{\log n}{\log(n\hat{p})}\right) \).
Below the connectivity threshold

**Theorem**

If \( \frac{1}{n} \ll \hat{p} \leq \frac{c \log n}{n} \) with \( c < 1 \), then the reachability threshold from any \( s \in [n] \) in \( \mathcal{M}_{n,p,q} \) is, a.a.s., equal to \( \Theta\left(\frac{\log n}{np}\right) \). Moreover, the optimal time of flooding from any \( s \in [n] \) in \( \mathcal{M}_{n,p,q} \) is, a.a.s., equal to

\[
\Theta\left(\frac{\log n}{np} + \frac{\log n}{\log np}\right),
\]

and the \( k \)-active flooding protocol from any \( s \in [n] \) in \( \mathcal{M}_{n,p,q} \) with \( k = \Omega\left(\frac{\log n}{np}\right) \) completes a.a.s. in optimal time.

**Theorem**

If \( 0 < \hat{p} \leq \frac{c}{n} \) for some constant \( c > 0 \), then, a.a.s., both the reachability threshold and the optimal flooding time from any \( s \in [n] \) in \( \mathcal{M}_{n,p,q} \) are equal to \( \Theta\left(\frac{\log n}{np}\right) \).
Proof

Two cases:
- $G_{n, \hat{p}}$ is likely to have a giant component: $\hat{p} \geq c/n$ for some constant $c > 1$
- $G_{n, \hat{p}}$ is not likely to have a giant component: $\hat{p} < 1/n$

1. Increase $k$ so that $G_{n, \hat{p}, p}^{(k)}$ is likely to have a giant component
2. connect external nodes to the giant component
Conclusion

There is a need for a generic model capturing network dynamism

Such generic model should capture:

- Time dependencies
- Spatial dependencies