AATG'20

Simple Graph Dynamics with Churn

Andrea Clementi[♦]

joint work with

L. Becchetti[♥], F. Pasquale[♦], L. Trevisan[♣], and I. Ziccardi[♠]

[♥]Sapienza Università di Roma, [♦]Tor Vergata Università di Roma,
[♣]Bocconi University, [♠]Universitá dell'Aquila

Our Research Activity since 2007 on Dynamic Graphs

General Goal: Study of Self-Organization in Population Systems

Our Research Activity since 2007 on Dynamic Graphs

General Goal: Study of Self-Organization in Population Systems

Local Interaction Rules in Population Systems:

Natural **Dynamics** = Simple Distributed Algorithms

Main Properties of **Dynamics**:

Homogenous: All agents run the same rule at every time

Local Communication: Few, short messages with few neighbors

Node Interactions: Opportunistic/random interactions among the nodes

Natural: See *"Natural Algorithms"* (Chazelle - Comm. ACM 2012)

A Fundamental Task: Network Formation and Maintenaince

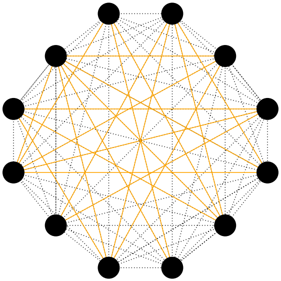
The Algorithmic Goal:

A finite set V of nodes (peers), interacting via a fixed communication graph H, wants to construct and keep a **dynamic subgraph** $\mathcal{G} = \{G_t = (N_t, E_t),\ t \geq 0\}$ of H such that:

- At every time $t \geq 1$, G_t is **sparse**
- At every time $t \geq 1$, G_t has **good connectivity properties** (with high probability, i.e., *w.h.p.*) and/or **Information Spreading** over $\mathcal{G}$ is **Fast**

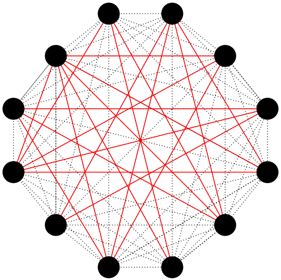
Our Research Activity on Graph Dynamics

Figure: Distributed Graph Sparsification: Connection Requests



Our Research Activity on Graph Dynamics

Figure: Distributed Graph Sparsification: Sparse Spanning Subgraph



Our Research Activity on Graph Dynamics in 2019

Network Formation and Maintenance via Natural Graph Dynamics

Crucial Model Assumption: fixed, time-invariant set V of nodes

- Our paper in ACM-SIAM SODA'20 (Francesco Pasquale's Talk at AATG'19)
- Our paper in ACM SPAA'20

Our Research Activity for 2020

Network Formation and Maintenance via Graph Dynamics

- **New Challenging Issue: Introducing Node Churn**

Our Research Activity for 2020

Network Formation and Maintenance via Graph Dynamics

- **New Challenging Issue: Introducing Node Churn**

Technical Question:

- **Consider a Graph Dynamics in the presence of Node Churn that yields a sparse dynamic graph and analyze its Connectivity Properties and Information Spreading**

A Key Connectivity Property: Vertex Expansion

Outer boundary

Let $G = (V, E)$ be a graph of n nodes. For each $S \subseteq V$, $\partial_{out}(S)$ is the *outer boundary of S*, i.e. the set of nodes in $V - S$ with at least one neighbor in S.

A Key Connectivity Property: Vertex Expansion

Outer boundary

Let $G = (V, E)$ be a graph of n nodes. For each $S \subseteq V$, $\partial_{out}(S)$ is the *outer boundary of* S, i.e. the set of nodes in $V - S$ with at least one neighbor in S.

Vertex isoperimetric number

The *vertex isoperimetric number* is

$$h_{out}(G) = \min_{0 \leq |S| \leq n/2} \frac{|\partial_{out}(S)|}{|S|} \tag{1}$$

A Key Connectivity Property: Vertex Expansion

Outer boundary

Let $G = (V, E)$ be a graph of n nodes. For each $S \subseteq V$, $\partial_{out}(S)$ is the *outer boundary of* S, i.e. the set of nodes in $V - S$ with at least one neighbor in S.

Vertex isoperimetric number

The *vertex isoperimetric number* is

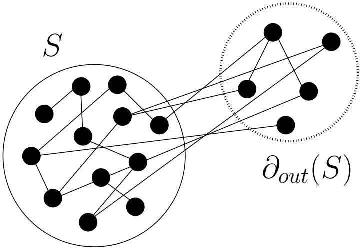
$$h_{out}(G) = \min_{0 \leq |S| \leq n/2} \frac{|\partial_{out}(S)|}{|S|} \tag{1}$$

Vertex expansion

Let $\varepsilon > 0$ be an arbitrary constant. Then, G is a ε-*expander* if $h_{out}(G) \geq \varepsilon$.

A Key Connectivity Property: Vertex Expansion

Figure: The Vertex Expansion of a Subset of Vertices



A Key Epidemic Process: Flooding

The Flooding Process

Consider a dynamic graph $\mathcal{G} = \{G_t = (N_t, E_t),\ t \geq 0\}$. Let s be the (first) infected node joining the graph at round t_0 and let $I_0 = \{s\} \subseteq V_{t_0}$
Then, at each round $t \geq t_0$, the *Flooding Process* is defined by the following sequence of subsets of *infected* nodes:

$$I_t = \left(I_{t-1} \bigcup I'_t \right) \bigcap V_t \,, \ \text{ where } \ I'_t = \{ v \in N_{t-1} | \exists u \in I_{t-1} : (u,v) \in E_{t-1} \}$$

A Key Epidemic Process: Flooding

The Flooding Process

Consider a dynamic graph $\mathcal{G} = \{G_t = (N_t, E_t),\ t \geq 0\}$. Let s be the (first) infected node joining the graph at round t_0 and let $I_0 = \{s\} \subseteq V_{t_0}$
Then, at each round $t \geq t_0$, the *Flooding Process* is defined by the following sequence of subsets of *infected* nodes:

$$I_t = \left(I_{t-1} \bigcup I'_t\right) \bigcap V_t \,, \text{ where } I'_t = \{v \in N_{t-1} | \exists u \in I_{t-1} : (u, v) \in E_{t-1}\}$$

Remark.

In the case of *Static* Graphs:

$$\text{Flooding Time} \quad = \quad \text{Diameter}$$

Network Formation and Maintenance with Node Churn

Previous Analytical Work

- Dynamic-Graph Protocols with access to Central Servers and/or Random Oracles: [Pandurangan et al. - IEEE FOCS'03], [Duchon et al. - LATIN'14]

Network Formation and Maintenance with Node Churn

Previous Analytical Work

- Dynamic-Graph Protocols with access to Central Servers and/or Random Oracles: [Pandurangan et al. - IEEE FOCS'03], [Duchon et al. - LATIN'14]
- Dynamic-Graph Protocols based on Random Walks: [Cooper et Al - Combinatorics, Probability and Computing 2007], [Law and Siu - IEEE INFOCOM'03], [Augustine et al - IEEE FOCS'15]

SHARED FEATURE of Previous Work: NO NATURAL DYNAMICS

Protocols are **carefully designed** to get the desired properties

Our Contribution: The Starting Point

The Static Framework: No node churn; No edge changes

The simplest fully-random Graph Dynamics over the complete communication graph:

Our Contribution: The Starting Point

The Static Framework: No node churn; No edge changes

The simplest fully-random Graph Dynamics over the complete communication graph:

The d-Random Choice Protocol

- **Time** $t = 0$: a set of n nodes/ agents $V_0 = V$; an empty edge set $E_0 = \emptyset$.
- **Time** $t = 1$ $V_t := V$; Each node u selects **independently, u.a.r.** d *(out-)neighbor*s from V and connects to each of them. Add each selected link to $E_t = E$

Random Oracle

The d-**Random Choice Protocol** requires a simple PULL mechanism that each node can call to select one random node in the graph.

Our "Static" Starting Model

THEOREM (Popular Result :):))

For sufficiently large n, for any $d \geq 3$, at every step $t \geq 1$, the random graph $G_t(V_t, E_t)$ is a $\Theta(1)$-Expander, *with high probability* (for short, *w.h.p.*).

Our "Static" Starting Model

THEOREM (Popular Result :):))

For sufficiently large n, for any $d \geq 3$, at every step $t \geq 1$, the random graph $G_t(V_t, E_t)$ is a $\Theta(1)$-Expander, *with high probability* (for short, *w.h.p.*).

COROLLARY

The diameter of G and, so, its Flooding Time is $O(\log n)$, w.h.p..

Our Basic Dynamic Model: Informal Definition

Node Churn via (deterministic) Streaming

We adapt the d-**Random Choice Dynamics** to the simplest and **unrealistic** dynamic-graph model with **Node Churn**:

- nodes join/leave the network according to a discrete-time **streaming** process.
- edges of the leaving node disappear; active nodes replace their dying edges

Remark

Our Streaming Model is **unrealistic** , however,....
it allows to investigate Key Technical Issues that surely appear in more realistic and complex models.

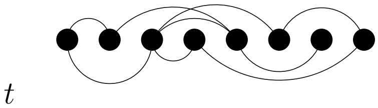
Our Streaming Model: Definition

A *Streaming Dynamic Graph with edge Regeneration* SDGR $\mathcal{G}(n, d)$ is a stochastic process $\{G_t = (N_t, E_t), t \geq 1\}$ defined as follows.

- **Node Churn Events.** $N_0 = \emptyset$. At each round $t \geq 1$, a new node joins N_t and *it stays alive* up to round $t + n$, then it leaves the game. So, at every $t \geq n$, the *oldest* node v leaves the network and a *new node u* joins it, i.e., $N_t := (N_{t-1} \setminus \{v\}) \cup \{u\}$.
- **Topology: The d-Random Choice Dynamics.** E_t evolves as follows:
  i) All the edges incident to the **leaving** node v disappear.
  ii) The new node u selects *independently, u.a.r.* d (out-)neighbors from N_t.
  iii) The nodes in N_t that lose some of their d (out-)edges (since v died), send new requests (independently, u.a.r from N_t) to keep (out-)degree d.

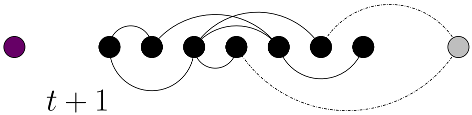
Our Streaming Model: SDGR $\mathcal{G}(n, d)$

Figure: Streaming Model



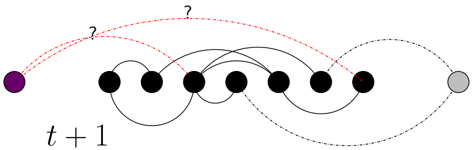
Our Streaming Model: SDGR $\mathcal{G}(n, d)$

Figure: Streaming Model



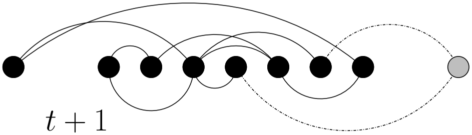
Our Streaming Model: SDGR $\mathcal{G}(n, d)$

Figure: Streaming Model



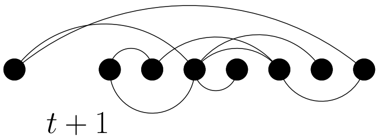
Our Streaming Model: SDGR $\mathcal{G}(n, d)$

Figure: Streaming Model



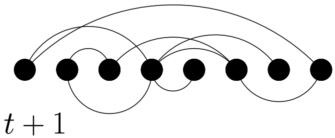
Our Streaming Model: SDGR $\mathcal{G}(n, d)$

Figure: Streaming Model



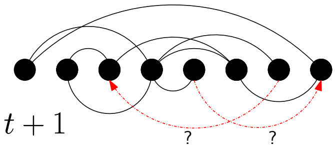
Our Streaming Model: SDGR $\mathcal{G}(n, d)$

Figure: Streaming Model



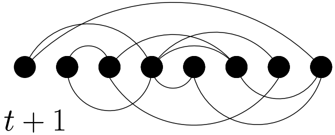
Our Streaming Model: SDGR $\mathcal{G}(n, d)$

Figure: Streaming Model



Our Streaming Model: SDGR $\mathcal{G}(n, d)$

Figure: Streaming Model



OUR CONTRIBUTION I: Vertex Expansion

(Main) THEOREM 1.

- **Streaming Model SDGR $\mathcal{G}(n, d)$.** For any sufficiently large d (i.e. $d \geq 14$), and for any $t \geq \Omega(n)$, the snapshot $G_t(N_t, E_t)$ is a $(1/10)$-expander, with probability $1 - 1/n^{\Theta(d)}$.

OUR CONTRIBUTION II: Flooding Time

The Flooding Process in the Streaming Model.

Consider a SDGR $\mathcal{G}(n, d) = \{G_t = (N_t, E_t),\ t \geq 0\}$. Let s be the infected node joining the graph at round t_0 and let $I_0 = \{s\} \subseteq V_{t_0}$
Then, at each round $t \geq t_0$, after applying the d-**Random Choice Dynamics**, attach the *Epidemic Process* defined by **Flooding**, i.e., by the time sequence of subsets of *infected* nodes:

$$I_t = \left(I_{t-1} \bigcup I'_t \right) \bigcap V_t \, , \ \text{ where } \ I'_t = \{ v \in N_{t-1} | \exists u \in I_{t-1} : (u, v) \in E_{t-1} \}$$

Flooding Time

(Main) THEOREM 2.

- **Streaming Model SDGR $\mathcal{G}(n, d)$.** For any sufficiently large d (i.e. $d \geq 14$), and for any $t \geq \Omega(n)$. Then, if an infected node is inserted at time step t, after $O(\log n)$ time steps, all nodes of the network will be infected, w.h.p.

Highlights of the Proof of THEOREM I : Vertex Expansion

Expansion of $G_t = (N_t, E_t)$

- **Main Technical Issue.** The different life times of the nodes in N_t make **correlation** among edges in E_t and a **non uniform edge probability**

Highlights of the Proof of THEOREM I : Vertex Expansion

Expansion of $G_t = (N_t, E_t)$

- **Main Technical Issue.** The different life times of the nodes in N_t make **correlation** among edges in E_t and a **non uniform edge probability**
- A good Intuition:
  Edges incident to **older** *nodes have more chances to belong to* E_t

Highlights of the Proof of THEOREM I : Vertex Expansion

Expansion of $G_t = (N_t, E_t)$

- **Main Technical Issue.** The different life times of the nodes in N_t make **correlation** among edges in E_t and a **non uniform edge probability**
- A good Intuition:
  Edges incident to **older** *nodes have* <u>*more*</u> *chances to belong to* E_t
- Ok, ...but how large can this **probability-gap** be ?

Highlights of the Proof of THEOREM I : Vertex Expansion

Expansion of $G_t = (N_t, E_t)$

- **LEMMA 1.** Let $k \leq t-1$ and let u be the node having age $k+1$. Then, if another node v in N_t is born before u, the probability that a **single request** of u has destination v is

$$\frac{\mathbf{1}}{\mathbf{n-1}}\left(\mathbf{1}+\frac{\mathbf{1}}{\mathbf{n-1}}\right)^{\mathbf{k}}, \qquad (2)$$

while, if v is born after u, the probability that a single slot of u has destination v is always $\leq \frac{\mathbf{1}}{\mathbf{n-1}}$

- **Good News.** Since $k \leq n$, Eq. (2) is $\leq \boldsymbol{\Theta}(\mathbf{1}/\mathbf{n})$

Highlights of the Proof of THEOREM I : Vertex Expansion

THEOREM

Let n be sufficiently large and $d \geq 21$. Then, for any $t \geq n$, the snapshot G_t of a SDGR $\mathcal{G}(n, d)$ is a vertex expander with parameter $\varepsilon \geq 0.1$, w.h.p.

Proof Strategy

We split the analysis in two cases:
Case 1. Small subsets, i.e., $|S| \leq n/4$,
Case 2. Large subsets, i.e., $n/4 \leq |S| \leq n/2$,

Remark

In both cases, the S expansion is obtained by only looking at the **out-going** edges of set S, i.e., those edges determined by the d random slots of each node of S.

Highlights of the Proof of THEOREM I : Vertex Expansion

LEMMA (Case 1)

For every pair of vertex subsets (S, T) with $|S| \leq n/4$ and $|T| = 0.1|S|$, such that $S \cap T = \emptyset$, the event "*all the out-neighbors of S are in T*", i.e. $\partial_{out}(S) \subseteq T$, does happen with negligible probability, i.e., with probability $O(1/n^{\Theta(1)})$.

Highlights of the Proof of THEOREM I : Vertex Expansion

LEMMA (Case 1)

For every pair of vertex subsets (S, T) with $|S| \leq n/4$ and $|T| = 0.1|S|$, such that $S \cap T = \emptyset$, the event "*all the out-neighbors of S are in T*", i.e. $\partial_{out}(S) \subseteq T$, does happen with negligible probability, i.e., with probability $O(1/n^{\Theta(1)})$.

Proof

For any S and any $T \subseteq N_t - S$, we define the event $A_{S,T} = \{\partial_{out}(S) \subseteq T\}$ So, we have that

$$\mathbf{Pr}\left(\min_{n/4 \leq |S| \leq n/2} \frac{|\partial_{out}(S)|}{|S|} \leq 0.1\right) \leq \sum_{\substack{n/4 \leq |S| \leq n/2 \\ |T|=0.1|S|}} \mathbf{Pr}\left(A_{S,T}\right) . \tag{3}$$

The next step is to upper bound $\mathbf{Pr}\left(A_{S,T}\right)$.

Streaming Model SDGR Technical proofs

LEMMA (Case 1)

$\mathbf{Pr}\left(A_{S,T}\right)$ is upper bounded by the probability that each request of the nodes in S has destination in $S \cup T$.

From **Lemma 1** (the bound on the edge probability), since $k \leq n-1$, the probability that any request of u has destination any node v is at most $\mathbf{e}/(\mathbf{n}-\mathbf{1})$.

Since to have $\partial_{out}(S) \subseteq T$, each request of $u \in S$ must have destination in $S \cup T$, it holds

$$\mathbf{Pr}\left(A_{S,T}\right) \leq \left(\frac{\mathbf{e}}{\mathbf{n}-\mathbf{1}} \cdot |S \cup T|\right)^{d|S|} . \tag{4}$$

So, from (3) and (4), for any $d \geq 21$, and standard calculus,

$$\mathbf{Pr}\left(\min_{1\leq|S|\leq n/4} \frac{|\partial_{out}(S)|}{|S|} \leq 0.1\right) \leq \sum_{s=1}^{n/4} \binom{n}{s}\binom{n-s}{0.1s}\left(\frac{1.1s \cdot e}{n-1}\right)^{ds} \leq \frac{1}{n^4} . \tag{5}$$

Further Results I: The Poisson Dynamic model with edge Regeneration - PDGR

A PDGR $\mathcal{G}(\lambda, \mu, d)$ is a stochastic process $\{G_t = (N_t, E_t) \,:\, t \in \mathbb{R}^+\}$, where:

- **Node Churn Process [Pandurangan et al. - IEEE FOCS'03]**. Initially $N_0 = \emptyset$ and the node insertions in N_t are modelled by a sequential *Poisson process with mean* λ. Moreover, once a node is activated, its *life time* has *exponential distribution of parameter* μ.

- **Topology: The d-Random Choice Dynamics.**

Further Results I: The Poisson Dynamic model with edge Regeneration - PDGR

THEOREM (Poisson Model)

- **PDGR** $\mathcal{G}(\lambda, \mu, d)$ - **Expansion.** Let $\lambda = 1$ and $n = 1/\mu$, and let $d \geq 35$. Then, for any $t \geq \Omega(n \log n)$, the snapshot $G_t(N_t, E_t)$ is a $(1/10)$-expander, with probability $1 - 1/n^{\Theta(1)}$.
- **Poisson Model PDGR** $\mathcal{G}(\lambda, \mu, d)$ - **Flooding.** Let $\lambda = 1$ and $n = 1/\mu$, and let $d \geq 35$. Then, for any $t \geq \Omega(n \log n)$, if an infected node is inserted at time t, after $O(\log n)$ **flooding rounds**, all nodes of the network will be infected, w.h.p.

Further Results II: Dynamic Models with No Edge Regeneration

A “parsimonious” version of the d-**Random Choice Dynamics** over the Streaming and Poisson models with **no Edge Regeneration**.

Our Results:

- **Negative Results:**

  a There can be $\Theta(n)$ **isolated** nodes at every time

  b There is **constant** probability that a joining infected nodes fails to infect more than few nodes...

Further Results II: Dynamic Models with No Edge Regeneration

A “parsimonious” version of the d-**Random Choice Dynamics** over the Streaming and Poisson models with **no Edge Regeneration**.

Our Results:

- **Negative Results:**

  a There can be $\Theta(n)$ **isolated** nodes at every time

  b There is **constant** probability that a joining infected nodes fails to infect more than few nodes...

- **Positive Results:**

  a For $d = \Omega(1)$, at every time step t, every vertex subset of size $\geq n/10$, of the snapshot G_t has $\Theta(1)$-**expansion**, w.h.p..

  b For some $d = \Omega(1)$, a joining infected node infects $\mathbf{0.9} \cdot \mathbf{n}$ nodes within **log** time, with Prob. $\geq \mathbf{0.9}$

Open Question and the End

Major Open Question:

Design and Analysis of **Natural** Graph Dynamics in the presence of Node Churn that yield **Bounded-Degree Topologies** with good connectivity properties, w.h.p.

Open Question and the End

Major Open Question:

Design and Analysis of **Natural** Graph Dynamics in the presence of Node Churn that yield **Bounded-Degree Topologies** with good connectivity properties, w.h.p.

THANKS!!!!!!