Scalable and Reliable Searching in Unstructured Peer-to-Peer Systems

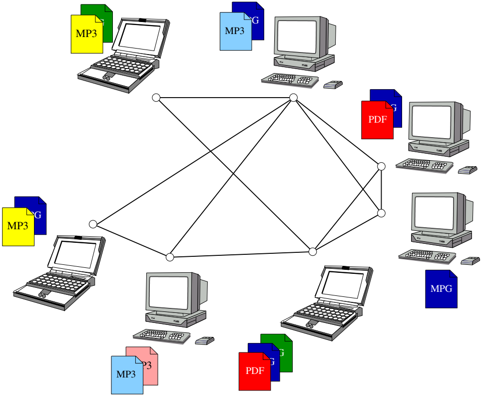
Stratis Ioannidis[1] Peter Marbach[2]

[1]Thomson
Paris, France

[2]University of Toronto
Toronto, ON. Canada

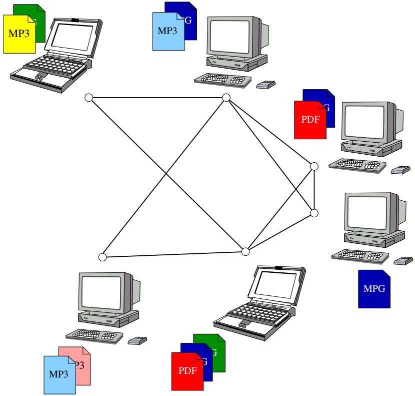
YEQT III
Eindhoven,
Nov. 21st, 2009

Motivation: Unstructured P2P File-Sharing Systems



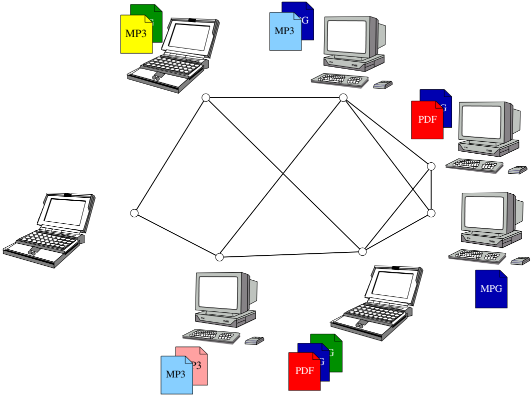
Peers form a network with the purpose of sharing files

Motivation: Unstructured P2P File-Sharing Systems



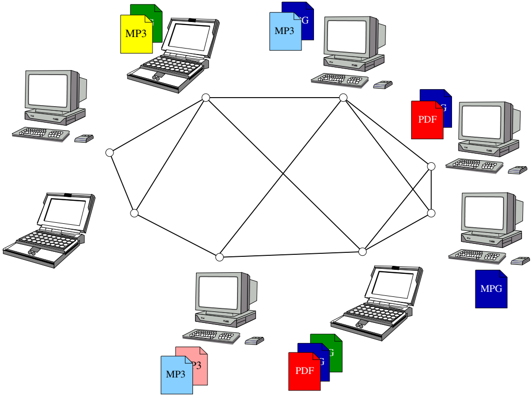
The system is dynamic

Motivation: Unstructured P2P File-Sharing Systems



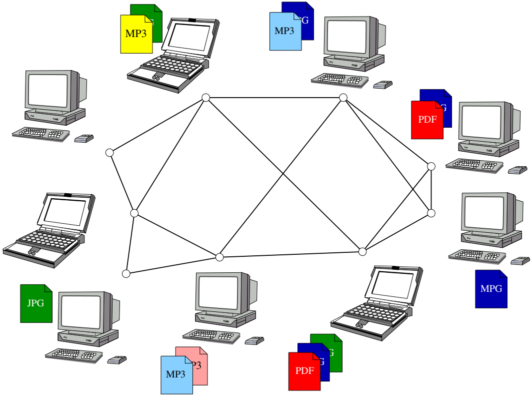
The system is dynamic

Motivation: Unstructured P2P File-Sharing Systems



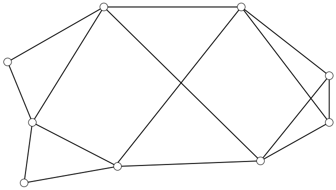
The system is dynamic

Motivation: Unstructured P2P File-Sharing Systems



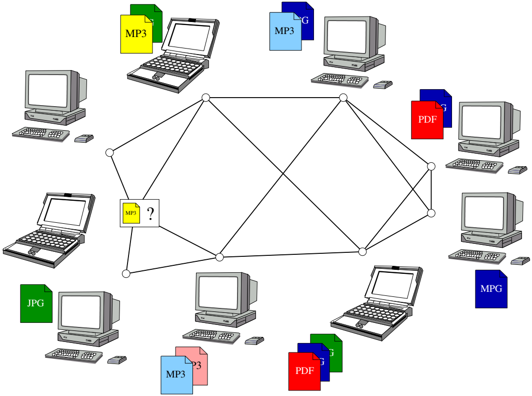
The system is dynamic

Motivation: Unstructured P2P File-Sharing Systems



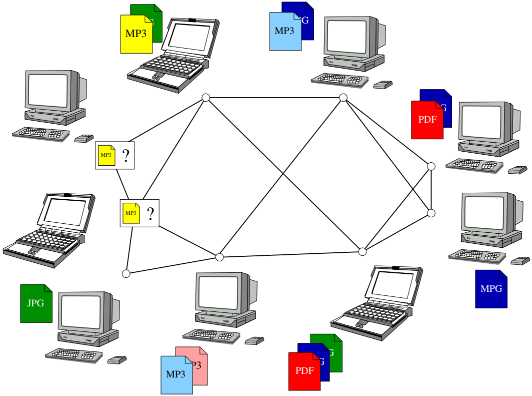
Unstructured: Overlay graph may be arbitrary

Motivation: Unstructured P2P File-Sharing Systems



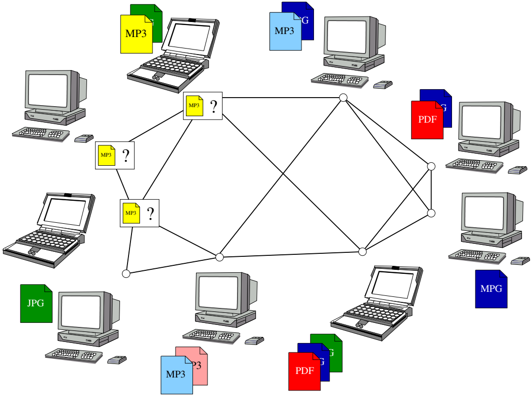
Peers propagate queries over the p2p network

Motivation: Unstructured P2P File-Sharing Systems



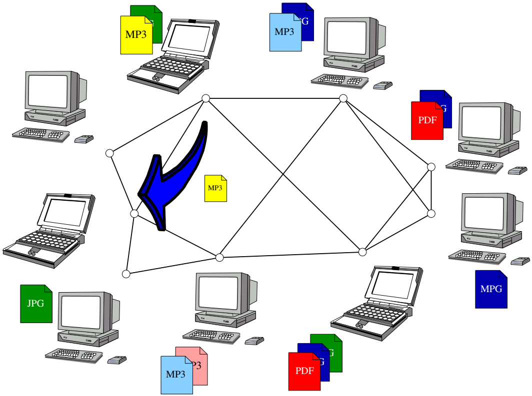
Peers propagate queries over the p2p network

Motivation: Unstructured P2P File-Sharing Systems



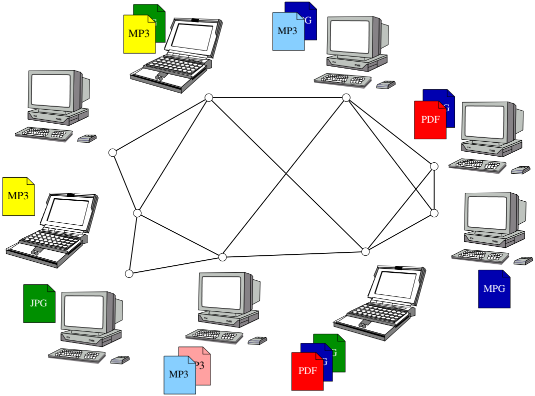
Peers propagate queries over the p2p network

Motivation: Unstructured P2P File-Sharing Systems



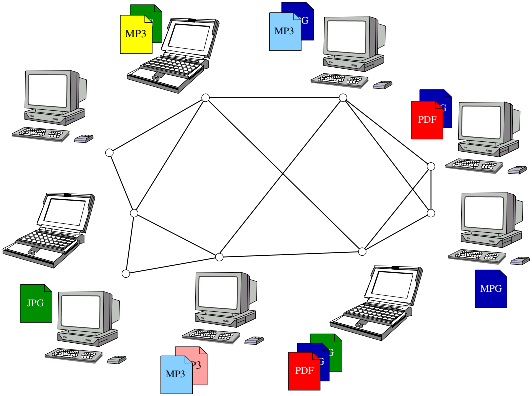
Peers respond by providing requested file

Motivation: Unstructured P2P File-Sharing Systems



Peers respond by providing requested file

Motivation: Unstructured P2P File-Sharing Systems



If the query fails, the peer does not retrieve the file

Goal

Query propagation (search) mechanisms that are both

- scalable and
- reliable.

Goal

Query propagation (search) mechanisms that are both

- scalable and
- reliable.

Challenges: Cope with

- Arbitrary topology
- Churn

Related Work

Proposed Search Mechanisms

- Random Walk [Lv et al., 2002, Gkantsidis et al., 2004]
- Expanding Ring [Tewari and Kleinrock, 2006, Lv et al., 2002]
- k-Parallel walks [Lv et al., 2002]
- Random walk with look-ahead [Gkantsidis et al., 2005, Puttaswamy et al., 2008]
- Budget-based forwarding [Terpstra et al., 2007, Gkantsidis et al., 2005]
- Proactive replication [Cohen and Shenker, 2002, Tewari and Kleinrock, 2006]
- ...

Related Work

Models:

- No overlay graph (Uniform sampling)
  [Lv et al., 2002, Cohen and Shenker, 2002, Terpstra et al., 2007]
- Static random graph (no churn)
  [Gkantsidis et al., 2004, Puttaswamy et al., 2008]
- Markovian graph models, but no search mechanisms
  [Law and Siu, 2003, Ganesh et al., 2007, Feder et al., 2006, Mahlmann and Schindelhauer, 2005]

Our Contributions

1. Markovian model that incorporates churn

Our Contributions

1. Markovian model that incorporates churn
2. We show that the random walk and the expanding ring mechanisms cannot be scalable and reliable!

Our Contributions

1. Markovian model that incorporates churn
2. We show that the random walk and the expanding ring mechanisms cannot be scalable and reliable!
3. We propose a mechanism that is both scalable and reliable.

Modelling Assumptions

Overlay graph: d-regular

Fixed size n:

File “popularity” $\neq$ File “availability”

Modelling Assumptions

Overlay graph: d-regular

- Peers try to maintain a constant number of connections.

Fixed size n:

File "popularity" $\neq$ File "availability"

Modelling Assumptions

Overlay graph: d-regular

- Peers try to maintain a constant number of connections.

Fixed size n:

- Long term growth (*e.g.* within months)
- Short term (*e.g.* day or week) size stability: Operating size n

File "popularity" $\neq$ File "availability"

Modelling Assumptions

Overlay graph: d-regular

- Peers try to maintain a constant number of connections.

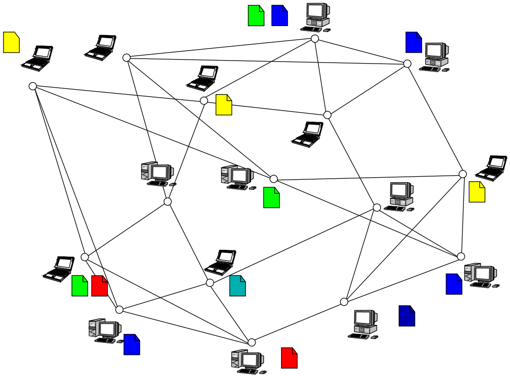
Fixed size n:

- Long term growth (*e.g.* within months)
- Short term (*e.g.* day or week) size stability: Operating size n

File "popularity" $\neq$ File "availability"

- A file might be requested often but rarely be in the system, and vice-versa

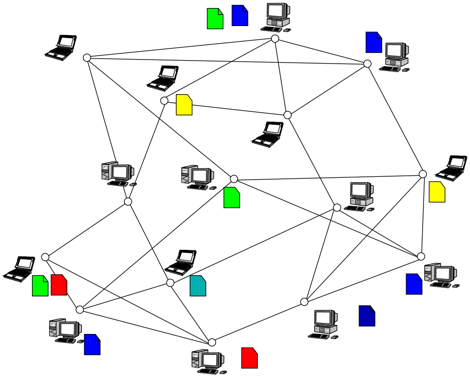
Churn Process



$\frac{1}{\mu}E_1$

Exponential lifetimes, mean $1/\mu$

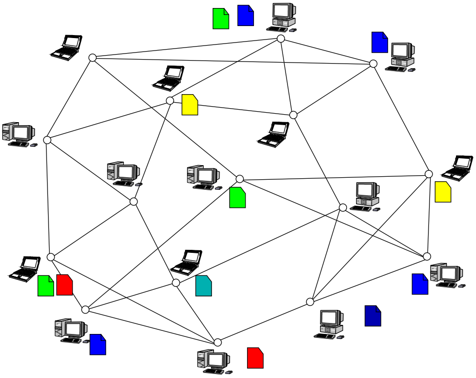
Churn Process



$$\frac{1}{\mu}E_1$$

Each departing peer is immediately replaced

Churn Process

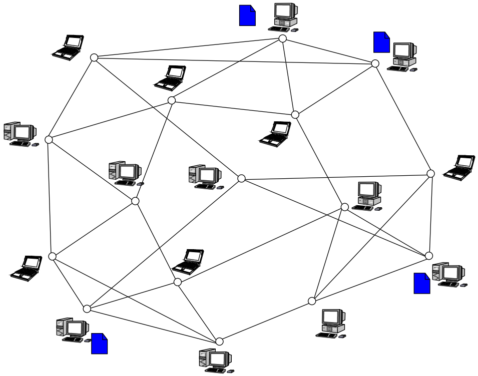


$\frac{1}{\mu}E_1$

system size: n

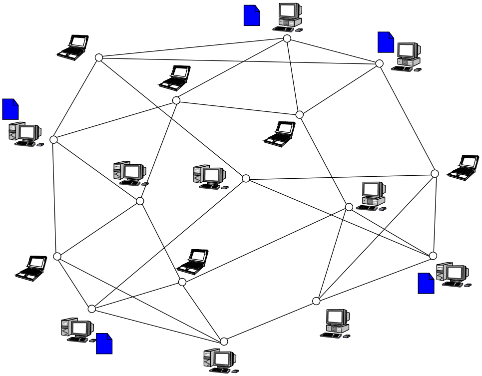
System size is fixed

File Request and File Publishing



Single file case.

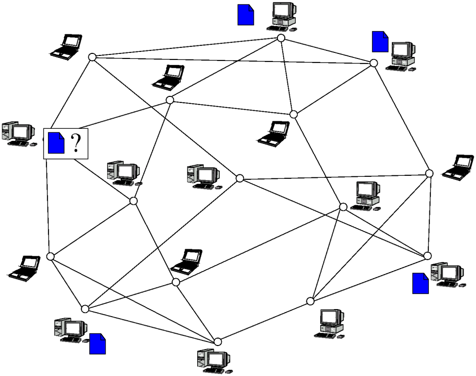
File Request and File Publishing



q_n

Incoming peer brings the file with probability q_n.

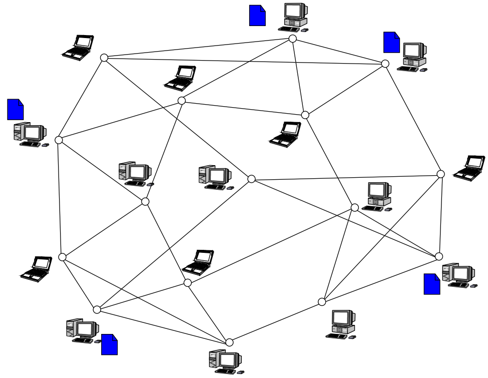
File Request and File Publishing



q_n, p_n

Incoming peer requests the file with probability p_n.

File Request and File Publishing

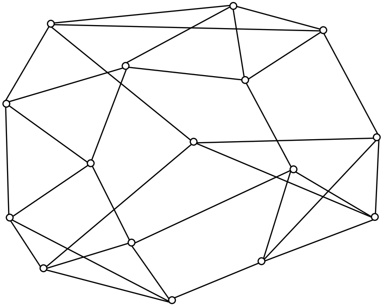


q_n, p_n

- $q_n = 0.01$
- $q_n \sim \frac{1}{n}$
- $q_n \sim \frac{1}{n^2}$

Expected number of peers bringing (requesting) file is nq_n (np_n).

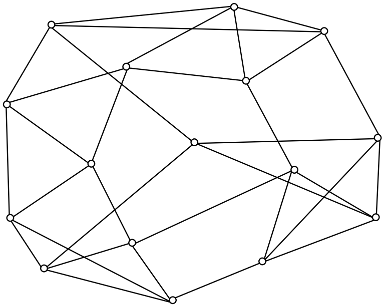
Overlay Graph



$$\{G(t)\}_{t\in\mathbb{N}}$$

$G(t)$: overlay graph at t-th departure/arrival epoch.

Overlay Graph

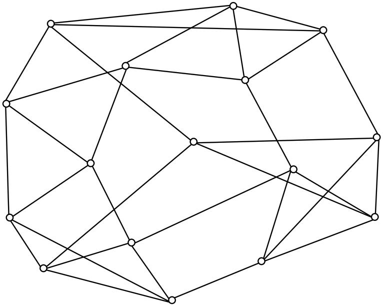


$$\{G(t)\}_{t\in\mathbb{N}}$$

$$G(t) \in \mathbb{G}_{n,d}$$

For all $t \geq 0, G(t)$ is d-regular graph with n vertices.

Overlay Graph

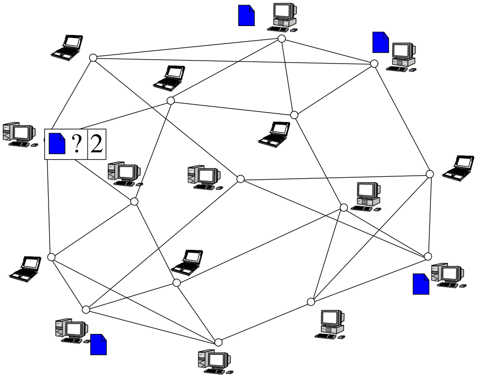


$$\{G(t)\}_{t\in\mathbb{N}}$$

$$G(t) \in \mathbb{G}_{n,d}$$

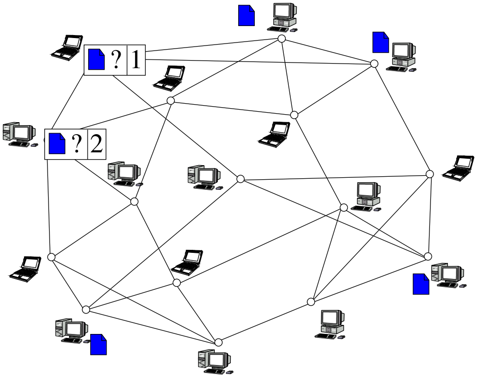
$\{G(t)\}_{t\in\mathbb{N}}$ is a Markov chain with state space $\mathbb{S}_{n,d} \subseteq \mathbb{G}_{n,d}$.

Random Walk with TTL_n



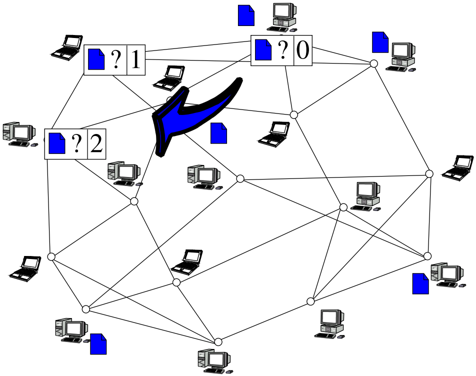
Query header initialized to TTL_n

Random Walk with TTL_n



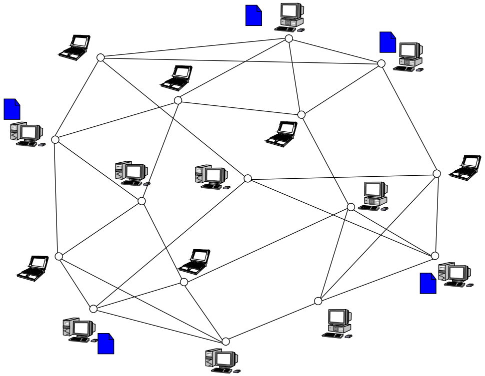
Header decremented with each hop

Random Walk with TTL_n



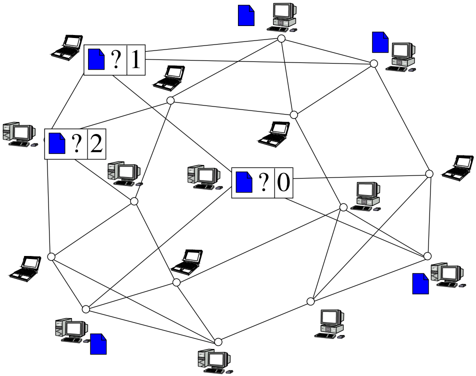
Query propagated until either file located or header is zero

Random Walk with TTL_n



Query propagated until either file located or header is zero

Random Walk with TTL_n



Query propagated until either file located or header is zero

Scalability and Reliability

Denote by

- ρ_n: the average traffic load per peer
- γ_n: the query success rate.

Scalability and Reliability

Definition

We will say that a search mechanism is *scalable* if,

$$\rho_n = O(1),$$

for all p_n,q_n.

I.e., the average load per peer ρ_n stays bounded as the system size n increases.

Scalability and Reliability

Definition

We will say that a search mechanism is *reliable* if

$$\text{if } q_n = \omega\left(\frac{1}{n}\right) \text{ then } \lim_{n\to\infty} \gamma_n = 1,$$

for all p_n.

I.e., if $\omega(1)$ peers bring the file, in expectation, almost all queries are guaranteed to succeed (asymptotically).

Random Walk with TTL_n

Theorem

The random walk mechanism cannot be both scalable and reliable.

Random Walk with TTL_n

Theorem

The random walk mechanism cannot be both scalable and reliable.

Intuition:

- If $\mathrm{TTL}_n = \omega(1)$, then queries for files not in system $\left(q_n = o\left(\frac{1}{n}\right)\right)$ generate an unbounded load.
- If $\mathrm{TTL}_n = O(1)$, then $\gamma_n \not\to 1$, even for files brought very often in the system $\left(q_n = \omega\left(\frac{1}{n}\right)\right)$.

Random Walk with TTL_n

Theorem

The random walk mechanism cannot be both scalable and reliable.

Intuition:

- If $\mathrm{TTL}_n = \omega(1)$, then queries for files not in system ($q_n = o\left(\frac{1}{n}\right)$) generate an unbounded load.
- If $\mathrm{TTL}_n = O(1)$, then $\gamma_n \not\to 1$, even for files brought very often in the system ($q_n = \omega\left(\frac{1}{n}\right)$).

Same result holds for expanding ring.

Solution

Idea: Stop queries for files not in system, *without affecting queries for files that are in the system*

Solution

Idea: Stop queries for files not in system, *without affecting queries for files that are in the system*

Q: How to tell that a file is not in the system?

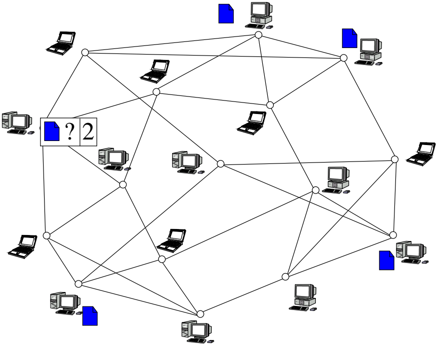
Solution

Idea: Stop queries for files not in system, *without affecting queries for files that are in the system*

Q: How to tell that a file is not in the system?

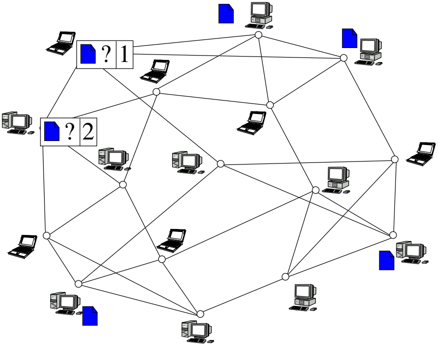
A: Use failed queries.

Absence of Evidence as Evidence of Absence



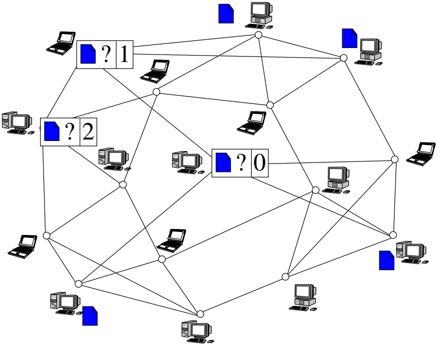
Suppose that a query fails to locate the file

Absence of Evidence as Evidence of Absence



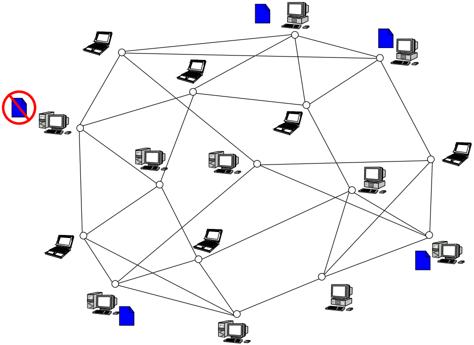
Suppose that a query fails to locate the file

Absence of Evidence as Evidence of Absence



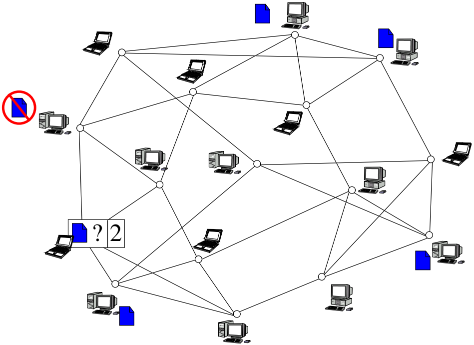
Suppose that a query fails to locate the file

Absence of Evidence as Evidence of Absence



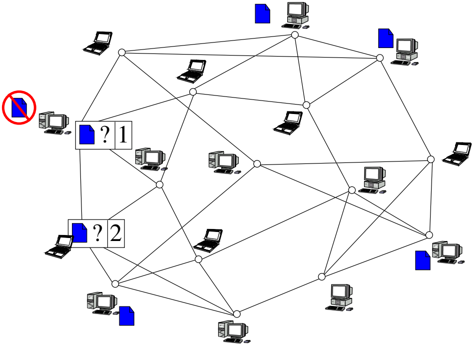
Store this information

Absence of Evidence as Evidence of Absence



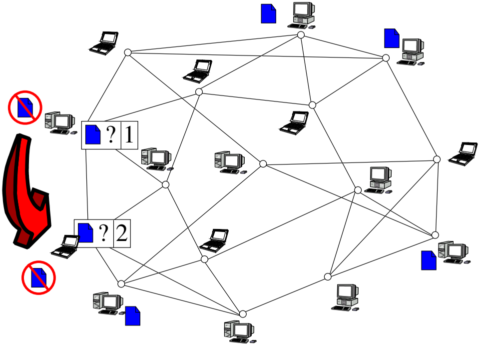
Use it to stop propagation of queries

Absence of Evidence as Evidence of Absence



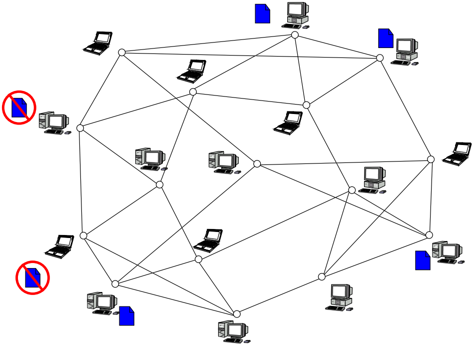
Use it to stop propagation of queries

Absence of Evidence as Evidence of Absence



Share it the same way as files

Absence of Evidence as Evidence of Absence



Random Walk using "Evidence of Absence"

Absence of Evidence as Evidence of Absence

What is the average traffic load per peer?
What about false negatives?

Scalability

Theorem

Assume that a graph sampled from the stationary distribution of $\{G(t)\}_{t\in\mathbb{N}}$ is an expander w.h.p. Then, the average traffic load per peer generated by a random walk with $\mathrm{TTL}_n = \Theta(n)$ that uses evidence of absence is

$$\rho_n = O(1),$$

i.e., it is bounded in n, irrespectively of p_n and q_n.

Scalability

Theorem

Assume that a graph sampled from the stationary distribution of $\{G(t)\}_{t\in\mathbb{N}}$ *is an expander w.h.p. Then, the average traffic load per peer generated by a random walk with* $\mathrm{TTL}_n = \Theta(n)$ *that uses evidence of absence is*

$$\rho_n = O(1),$$

i.e., it is bounded in n, irrespectively of p_n *and* q_n*.*

- If overlay is an expander *w.h.p.*, the random walk with EoA is scalable!

Scalability

Theorem

Assume that a graph sampled from the stationary distribution of $\{G(t)\}_{t\in\mathbb{N}}$ is an expander w.h.p. Then, the average traffic load per peer generated by a random walk with $\mathrm{TTL}_n = \Theta(n)$ that uses evidence of absence is

$$\rho_n = O(1),$$

i.e., it is bounded in n, irrespectively of p_n and q_n.

- If overlay is an expander *w.h.p.*, the random walk with EoA is scalable!
- Proved using bounds on hitting times of r.w. by Aldous and Fill.

Expander Graphs

Known Markov chains $\{G(t)\}_{t\in\mathbb{N}}$ with stationary distribution uniform over

- $\mathbb{MH}_{n,d}$: d-regular multi-graphs with a complete Hamiltonian decomposition
- $\mathbb{MI}_{n,d}$: d-regular multi-graphs with a 1-factorization

are expanders *w.h.p.*

Expander Graphs

Known Markov chains $\{G(t)\}_{t\in\mathbb{N}}$ with stationary distribution uniform over

- $\mathbb{MH}_{n,d}$: d-regular multi-graphs with a complete Hamiltonian decomposition
- $\mathbb{MI}_{n,d}$: d-regular multi-graphs with a 1-factorization

are expanders *w.h.p.*

Any distribution that is "almost uniform" over $\mathbb{G}_{n,d}$ will yield an expander

Reliability

Theorem

Assume that $\{G(t)\}_{t\in\mathbb{N}}$ are i.i.d., and that $G(t)$ is an expander w.h.p. Then, for the random walk with $\mathrm{TTL}_n = \Theta(n)$ that uses evidence of absence,

$$\text{if } q_n = \omega\left(\frac{1}{n}\right) \text{ then } \lim_{n\to\infty} \gamma_n = 1,$$

for all p_n.

- *I.e.*, the random walk using EoA is reliable!

Reliability

Theorem

Assume that $\{G(t)\}_{t\in\mathbb{N}}$ *are i.i.d., and that* $G(t)$ *is an expander w.h.p. Then, for the random walk with* $\mathrm{TTL}_n = \Theta(n)$ *that uses evidence of absence,*

$$\textit{if } q_n = \omega\left(\frac{1}{n}\right) \textit{ then } \lim_{n\to\infty} \gamma_n = 1,$$

for all p_n.

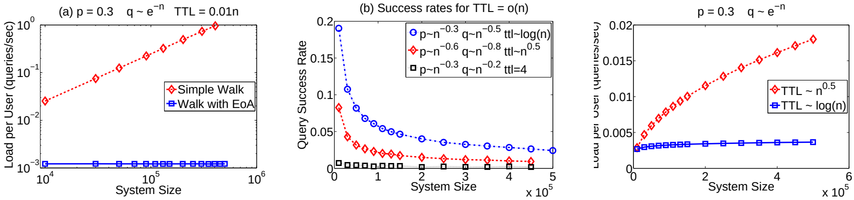
- *I.e.*, the random walk using EoA is reliable!
- Proved using fluid limit method by Benaïm and Le Boudec [2008].

Simulation Setup

- Law and Siu [2003] peer-to-peer system (Markov Chain over $\mathbb{MH}_{n,d}$).
- $\frac{1}{\mu} = 20\text{min}$.
- Arrival rate $n \cdot \mu$, $n = 10$ thousand to half a million.
- Degree 16.

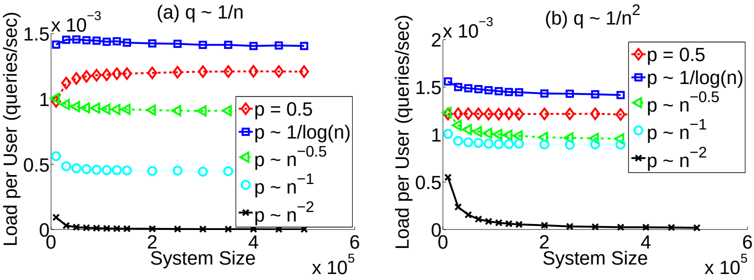
Random Walk Without Evidence of Absense

Traffic load and success rate of (traditional) random walk with TTL_n



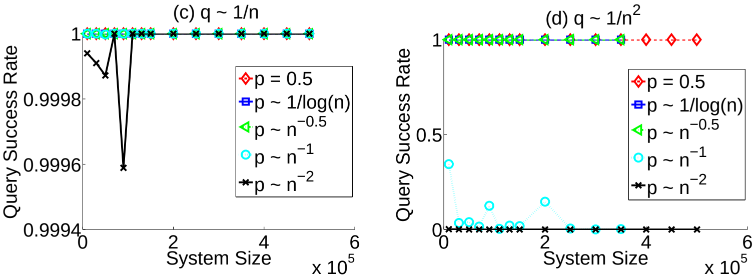
Random Walk Using Evidence of Absence -I

Traffic loads for data items brought in the system with publishing probabilities $q_n = 1000/n$ and $q_n = 1000^2/n^2$.



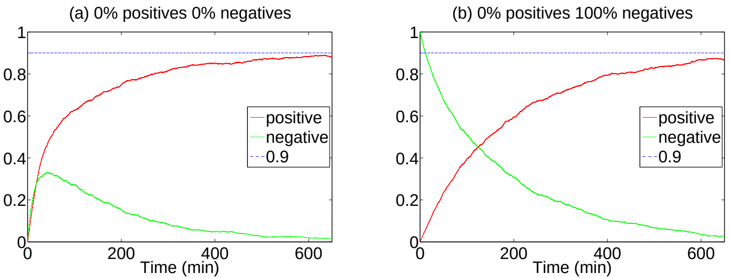
Random Walk Using Evidence of Absence -II

Success rates for data items brought in the system with publishing probabilities $q_n = 1000/n$ and $q_n = 1000^2/n^2$.



System Dynamics

System evolution for $p = 0.8$, $q = 0.1$, $n = 10,000$



Conclusions and Future Work

- Simple, distributed mechanisms yield scalability and reliability

Conclusions and Future Work

- Simple, distributed mechanisms yield scalability and reliability
- More sophisticated mechanisms
  - *k*-Parallel walks
  - Budget-based forwarding
  - Proactive replication
  - ...

Conclusions and Future Work

- Simple, distributed mechanisms yield scalability and reliability
- More sophisticated mechanisms
  - k-Parallel walks
  - Budget-based forwarding
  - Proactive replication
  - ...
- Different modelling assumptions:
  - What if overlay graph not an expander?

Conclusions and Future Work

- Simple, distributed mechanisms yield scalability and reliability
- More sophisticated mechanisms
  - k-Parallel walks
  - Budget-based forwarding
  - Proactive replication
  - ...
- Different modelling assumptions:
  - What if overlay graph not an expander?
- System dynamics vs. steady state behaviour

Thank You!

Case Study: Gnutella

Measurement studies:

- Ripeanu et al. [2002]
- Saroiu et al. [2002]
- Rasti et al. [2006]
- Li and Chen [2008]
- Stutzbach et al. [2008]
- Acosta and Chandra [2008]

Case Study: Gnutella Overlay

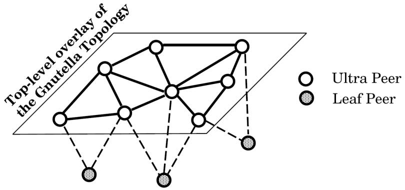


Figure source: Stutzbach et al. [2008]

- 2 tier-system (original version was flat)
- Ultra-peers know all the files shared by their leaves.
- Search happens on the ultra-peer level.

Case Study: Gnutella Overlay

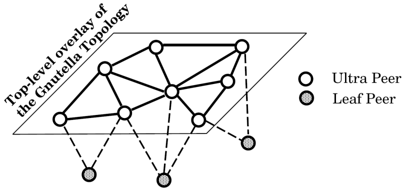


Figure source: Stutzbach et al. [2008]

- Ultra-peers connects to at most 32 other ultrapeers (Limewire, Bearshare).
- Each ultrapeer peer maintains at most 30 leaves in Limewire – 45 in Bearshare.
- Each leaf connects to at most 3 ultrapeers.

Case Study: Gnutella Overlay

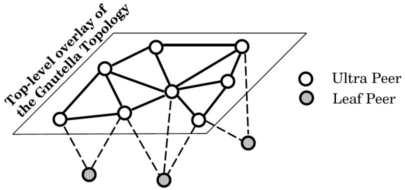


Figure source: Stutzbach et al. [2008]

Broken connections are replaced by cache, obtained by:

- Observing passing traffic
- Explicit cache exchanges with other users

Case Study: Gnutella Overlay

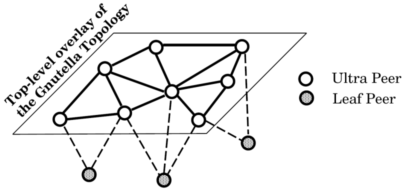


Figure source: Stutzbach et al. [2008]

Incoming peers use

- Caches from previous sessions
- Active probing
- Bootstrapping through a server or designated users.

Search Mechanisms

- Current implementation:
  - Constrained flooding over ultra-peers

Search Mechanisms

- Current implementation:
  - Constrained flooding over ultra-peers
- Methods proposed:

Search Mechanisms

- Current implementation:
  - Constrained flooding over ultra-peers
- Methods proposed:
  - Lv et al. [2002]:
    - Random walk
    - k-Random Walks
    - Expanding ring

Search Mechanisms

- Current implementation:
  - Constrained flooding over ultra-peers
- Methods proposed:
  - Lv et al. [2002]:
    - Random walk
    - k-Random Walks
    - Expanding ring
  - Gkantsidis et al. [2005]:
    - Random walk with lookahead.

Search Mechanisms

- Current implementation:
  - Constrained flooding over ultra-peers
- Methods proposed:
  - Lv et al. [2002]:
    - Random walk
    - *k*-Random Walks
    - Expanding ring
  - Gkantsidis et al. [2005]:
    - Random walk with lookahead.
  - Chawathe et al. [2003], Gkantsidis et al. [2005]:
    - Biased/adaptive search strategies.

Graph Properties: Growth, Oct 2004 - Jan 2006

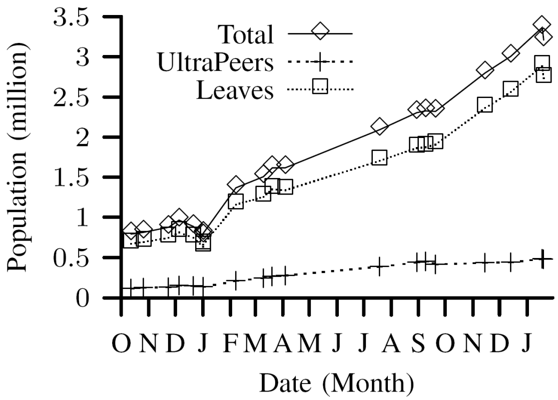
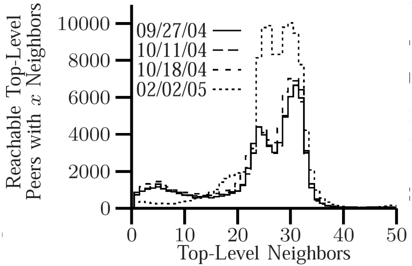


Figure source: Rasti et al. [2006]

Graph Properties: Degree Distribution



Crawl Date	Total Nodes	Ultra-Peers
09/27/04	725,120	110,208
10/11/04	779,535	116,967
10/18/04	806,948	120,229
02/02/05	1,031,471	158,345

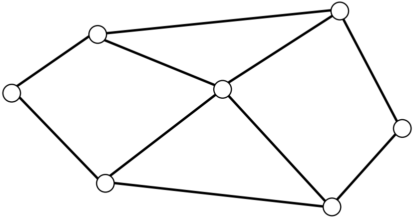
Figure and data source: Stutzbach et al. [2008].

Popularity $\neq$ Availability

Acosta and Chandra [2008]:

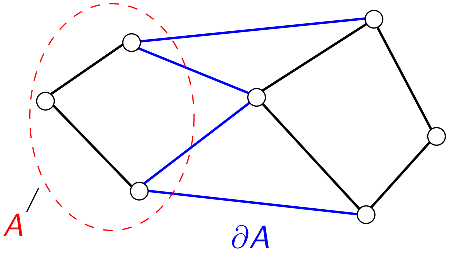
- There is no correlation betweent the popularity of a file and its availability in the system.
- 44.5% to 55.6% of queries cannot be matched to any file.

Edge Expansion Ratio



Let G be an undirected graph with vertex set V and edge set E.

Edge Expansion Ratio

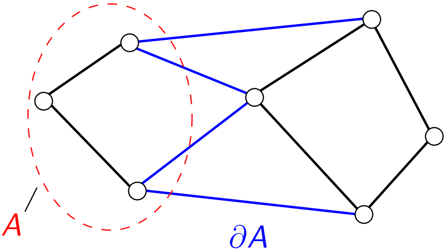


For $A \subset V$, the boundary of A is

$$\partial A = \{(i,j) \in E \mid i \in A \text{ and } j \in A^c\},$$

where $A^c = V \setminus A$

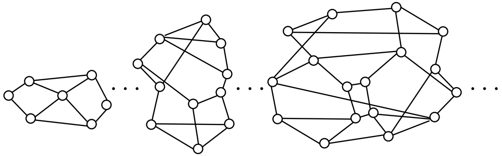
Edge Expansion Ratio



The edge expansion ratio h of G (Hoory et al. [2006]) is:

$$h = \min_{A \subset V, |A| \leq \frac{|V|}{2}} \frac{|\partial A|}{|A|}$$

Expander Graphs: Definition 1

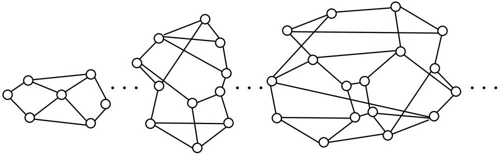


$G_7(V_7, E_7)$ $G_{11}(V_{11}, E_{11})$ $G_{17}(V_{17}, E_{17})$

$|V_n| = n$

Let $\{G_n\}_{n \geq n_0}$ be a sequence of graphs of increasing size, where G_n has size n.

Expander Graphs: Definition 1



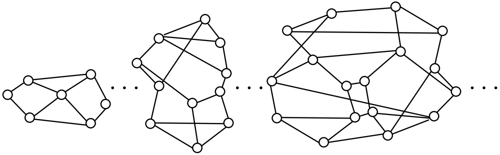
$G_7(V_7, E_7)$ $G_{11}(V_{11}, E_{11})$ $G_{17}(V_{17}, E_{17})$

$$|V_n| = n$$

$$d_{\max}(n) = \max_{i \in V_n} d_i \le d, \forall n$$

Assume that the graph sequence $\{G_n\}_{n \ge n_0}$ is of bounded degree.

Expander Graphs: Definition 1



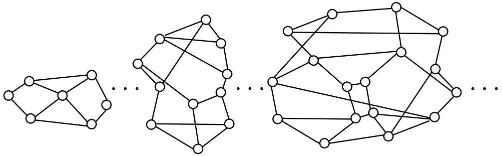
$G_7(V_7, E_7)$ $G_{11}(V_{11}, E_{11})$ $G_{17}(V_{17}, E_{17})$

$$|V_n| = n$$

$$d_{\max}(n) = \max_{i \in V_n} d_i \leq d, \forall n$$

Let $\{h_n\}_{n \geq n_0}$ be the corresponding sequence of expansion ratios.

Expander Graphs: Definition 1



$G_7(V_7, E_7)$ $G_{11}(V_{11}, E_{11})$ $G_{17}(V_{17}, E_{17})$

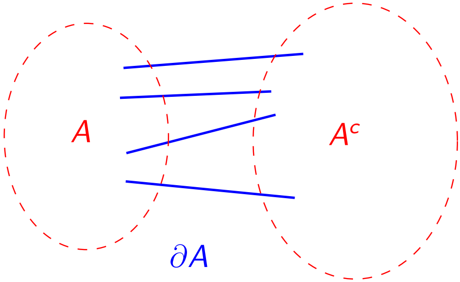
$$|V_n| = n$$

$$d_{\max}(n) = \max_{i \in V_n} d_i \leq d, \forall n$$

$$\exists\ \varepsilon > 0 \text{ such that, } \forall n \geq n_0,\ h_n \geq \varepsilon$$

Sequence $\{G_n\}_{n \geq n_0}$ is called an expander family if $\{h_n\}_{n \geq n_0}$ is bounded away from zero.

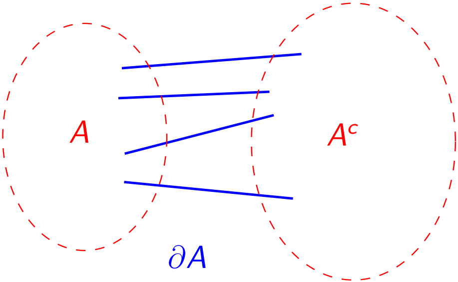
Intuition: Many Outgoing Edges



$$h_n|A| \leq |\partial A| \leq d|A|$$

for all sets $A \subset V_n$ with $|A| \leq |n|/2$.

Intuition: Many Outgoing Edges

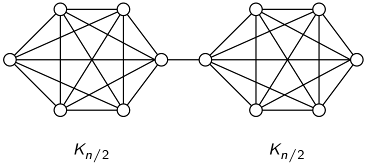


$$\epsilon|A| \leq |\partial A| \leq d|A|$$

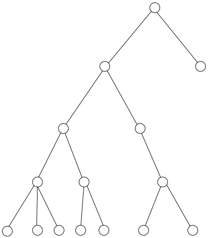
for all sets $A \subset V_n$ with $|A| \leq |n|/2$.

Non-examples

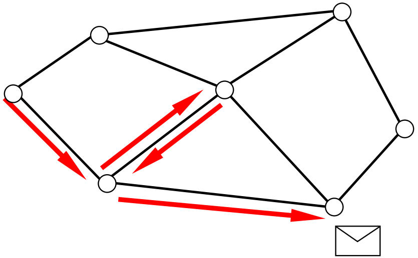
Barbell



Tree

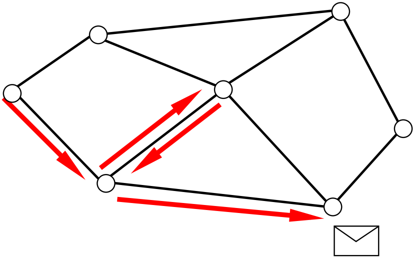


Expansion and Random Walks



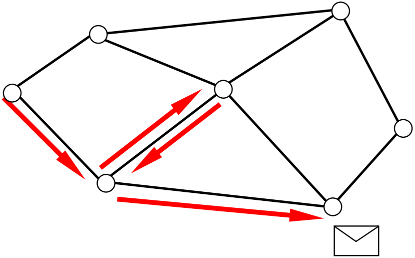
Random walk message propagation: forward a message to a neighbor chosen uniformly at random.

Expansion and Random Walks



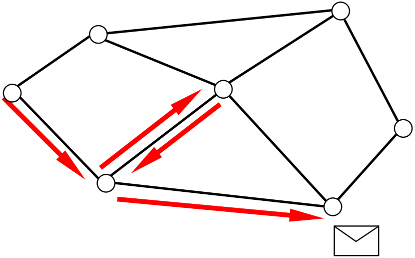
- Discrete time: each forwarding takes 1 time unit.
- Continuous time: each forwarding is exponentially distributed with mean 1 time unit.

Expansion and Random Walks



Let $X(t)$, $t \geq 0$, be the position of the message at time $t \geq 0$.

Expansion and Random Walks

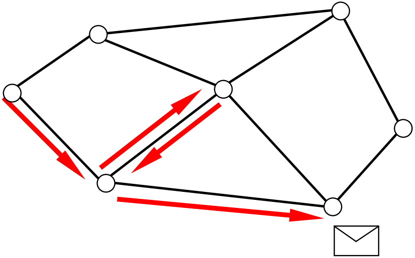


$X(t)$ is a Markov chain (Markov process in continuous time) with state space V and transition probabilities

$$P_{ij} = \begin{cases} \frac{1}{d_i}, & \text{if } i \text{ is connected to } j \\ 0, & \text{o.w.} \end{cases}$$

where d_i the degree of vertex i.

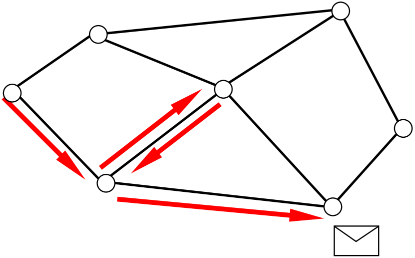
Expansion and Random Walks



$$\pi_j = \lim_{t \to \infty} \mathbf{P}_i(X(t) = j) = \frac{d_j}{\sum_k d_k} \text{ a.s.}$$

- Discrete time: G connected, non-bipartite.
- Continuous time: G connected.

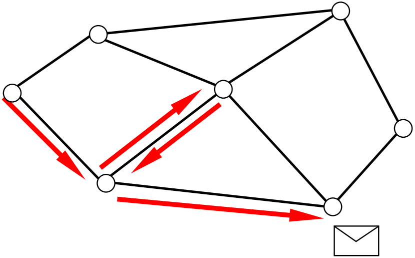
Expansion and Random Walks



If G is regular ($d_i = d$ for all i in V) then

$$\pi_j = \lim_{t\to\infty} \mathbf{P}_i(X(t) = j) = \frac{1}{|V|} \text{ a.s.}$$

Expansion and Random Walks



The relaxation time τ of G (Aldous and Fill) is

$$\tau = \frac{1}{1-\lambda_2}$$

where λ_2 the second largest eigenvalue of the transition probability matrix $[P_{ij}]$.

Expansion and Relaxation Time

The edge expansion ratio and the relaxation time are related as follows [Chung, 1997, Hoory et al., 2006]:

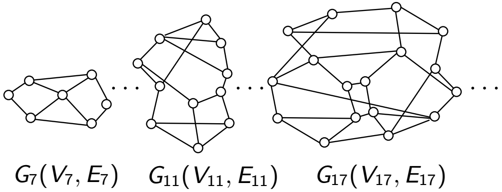
$$d_{\min}\frac{1}{2\tau} \leq h \leq d_{\max}\sqrt{\frac{2}{\tau}},$$

where

$$d_{\max} = \max_{i \in V} d_i, \qquad d_{\min} = \min_{i \in V} d_i$$

the maximum and minimum degrees of the graph, respectively.

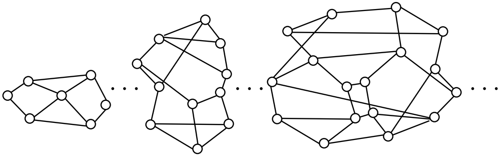
Expander Graphs: Definition 2



$|V_n| = n$

Let $\{G_n\}_{n \geq n_0}$ be a bounded-degree sequence, and $\{\tau_n\}_{n \geq n_0}$ the corresponding relaxation time sequence.

Expander Graphs: Definition 2



$G_7(V_7, E_7)$ $G_{11}(V_{11}, E_{11})$ $G_{17}(V_{17}, E_{17})$

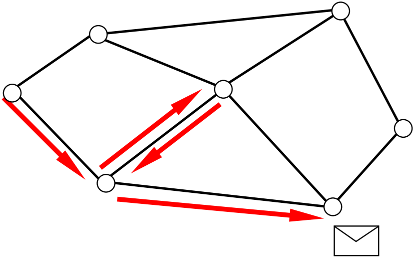
$$|V_n| = n$$

$$d_{\max}(n) = \max_{i \in V_n} d_i \leq d, \forall n$$

$$\exists\ M < \infty \text{ such that, } \forall n \geq n_0,\ \tau_n \leq M$$

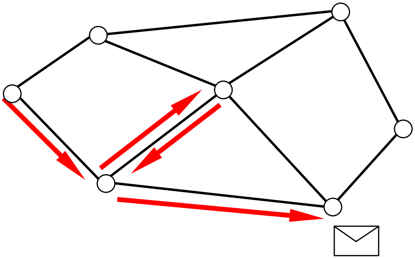
Sequence $\{G_n\}_{n \geq n_0}$ is an expander family iff $\{\tau_n\}_{n \geq n_0}$ is bounded.

Intuition 2: The Random Walk Mixes Fast



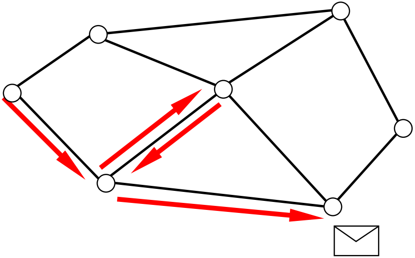
Consider the continuous-time random walk (jumps exponential with mean one).

Intuition 2: The Random Walk Mixes Fast



Denote with $\mathbf{P}_i(X(t) = j)$ the probability the random walk is at vertex j at time t, given that it started at vertex i.

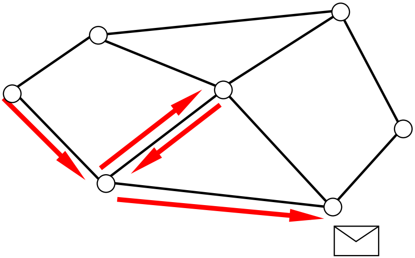
Intuition 2: The Random Walk Mixes Fast



The relaxation time relates to how fast the random walk converges to the steady state distribution.

$$d(t) = \inf_t \{t : \max_j |\mathbf{P}_i(X_t = j) - \pi_j| < \epsilon\} = O(\tau_n \log n)$$

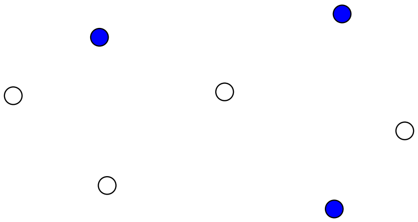
Intuition 2: The Random Walk Mixes Fast



The relaxation time relates to how fast the random walk converges to the steady state distribution.

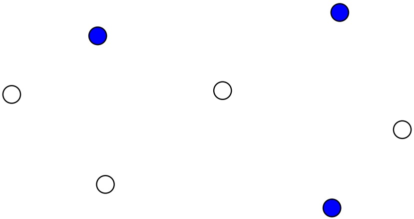
$$d(t) = \inf_t \{ t : \max_j |\mathbf{P}_i(X_t = j) - \pi_j| < \epsilon \} = O\left(M \log n \right)$$

Hitting Time (Aldous and Fill)



Let $A_n \subseteq V_n$ be a subset of V_n.

Hitting Time (Aldous and Fill)

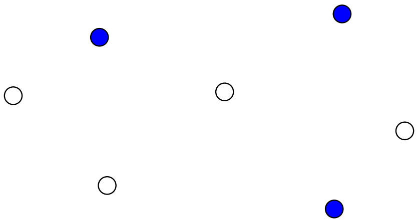


Let

$$T^u_{A_n} = \inf_t \{t : Y(t) \in A_n\}$$

be the time until an element in A_n is selected with uniform sampling.

Hitting Time (Aldous and Fill)

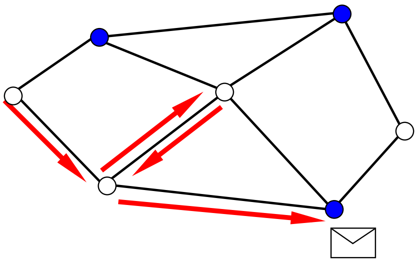


Let

$$T^u_{A_n} = \inf_t \{t : Y(t) \in A_n\}$$

be the time until an element in A_n is selected with uniform sampling.
Then, $\mathbb{E}[T^u_{A_n}] = \frac{n}{|A_n|}$

Hitting Time (Aldous and Fill)

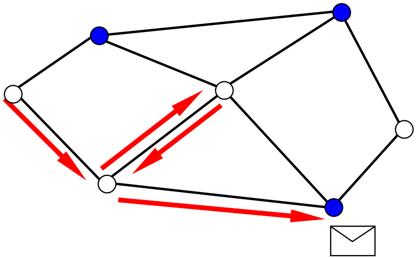


Let

$$T_{A_n} = \inf_t \{t : X(t) \in A_n\}$$

be the time it takes the random walk to hit set A_n.

Hitting Time (Aldous and Fill)

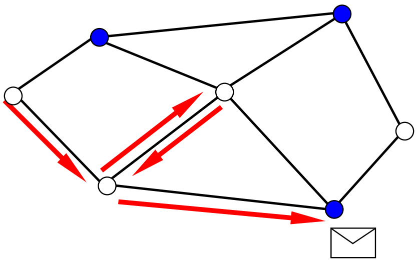


If the random walk starts uniformly outside A_n:

$$c^{-2}\frac{n}{|A_n|} - c^{-1} \leq \mathbb{E}_{u_{A_n{}^c}}[T_{A_n}] \leq c^2 \frac{\tau_n n}{|A_n|}$$

where $c = d_{\max}/d_{\min}$.

Hitting Time (Aldous and Fill)



If $\{G_n\}_{n \geq n_0}$ is an expander family then

$$\mathbb{E}_{u_{{A_n}^c}}[T_{A_n}] = \Theta\left(\frac{n}{|A_n|}\right) = \Theta\left(\mathbb{E}[T^u_{A_n}]\right)$$

Hitting Time (Aldous and Fill)

For continuous-time random walk, G_n regular:

$$\left(1 - \frac{2|A|\bar{\tau}_n}{n}\right) e^{-\frac{2|A|t}{n}} \leq \mathbf{P}_{u_{A^c}}(T_A > t) \leq e^{-\frac{|A|t}{n\bar{\tau}_n}}.$$

Definitions of *a.a.s.*,*w.h.p.*, and contiguity

Let ν_n be a probability measure over $\mathbb{G}_{n,d}$.

Definitions of *a.a.s.*,*w.h.p.*, and contiguity

Let ν_n be a probability measure over $\mathbb{G}_{n,d}$.

We say that $A_n \subseteq \mathbb{G}_{n,d}$ occurs *asymptotically almost surely* if

$$\lim_{n\to\infty} \nu_n(A_n) = 1.$$

Definitions of *a.a.s.*,*w.h.p.*, and contiguity

Let ν_n be a probability measure over $\mathbb{G}_{n,d}$.

We say that $A_n \subseteq \mathbb{G}_{n,d}$ occurs *asymptotically almost surely* if

$$\lim_{n \to \infty} \nu_n(A_n) = 1.$$

We say that $A_n \subseteq \mathbb{G}_{n,d}$ occurs *with high probability* if

$$\nu_n(A_n) = 1 - o\left(\frac{1}{n}\right).$$

Definitions of *a.a.s.*,*w.h.p.*, and contiguity

Let ν_n be a probability measure over $\mathbb{G}_{n,d}$.

We say that $A_n \subseteq \mathbb{G}_{n,d}$ occurs *asymptotically almost surely* if

$$\lim_{n\to\infty} \nu_n(A_n) = 1.$$

We say that $A_n \subseteq \mathbb{G}_{n,d}$ occurs *with high probability* if

$$\nu_n(A_n) = 1 - o\left(\frac{1}{n}\right).$$

We say that u_n is *contiguous* to ν_n if, for all $A_n \subseteq \mathbb{G}_{n,d}$,

$$\lim_{n\to\infty} u_n(A_n) = 1 \quad \text{iff} \quad \lim_{n\to\infty} \nu_n(A_n) = 1.$$

Random Regular Graphs are Expanders

Friedman [2003]: A random graph sampled uniformly from $\mathbb{G}_{n,d}$, $d \geq 3$, is an expander *a.a.s.*

Random Regular Graphs are Expanders

Friedman [2003]: A random graph sampled uniformly from $\mathbb{G}_{n,d}$, $d \geq 3$, is an expander *a.a.s.*

Definition

We will say that the system is unstructured if the stationary distribution of $\{G(t)\}_{t\in\mathbb{N}}$ is contiguous to the uniform distribution over $\mathbb{G}_{n,d}$.

…*i.e.*, it is "almost" uniform over all d-regular graphs.

Random Regular Graphs are Expanders

Friedman [2003]: A random graph sampled uniformly from $\mathbb{G}_{n,d}$, $d \geq 3$, is an expander *a.a.s.*

Definition

We will say that the system is unstructured if the stationary distribution of $\{G(t)\}_{t\in\mathbb{N}}$ is contiguous to the uniform distribution over $\mathbb{G}_{n,d}$.

. . . *i.e.*, it is "almost" uniform over all d-regular graphs.

Intuition: If the distribution is not almost uniform, then the overlay graph exhibits a certain structure.

Unstructured $\Rightarrow$ expander *a.a.s.*

Examples of "Almost"-Uniform Distributions

Wormald [1999]: Uniform distribution over

- $\mathbb{G}_{n,d}$: d-regular graphs
- $\mathbb{CG}_{n,d}$: connected d-regular graphs
- $\mathbb{H}_{n,d}$: d-regular graphs with a complete Hamiltonian decomposition
- $\mathbb{I}_{n,d}$: d-regular graphs with a 1-factorization

Examples of "Almost"-Uniform Distributions

Wormald [1999]: Uniform distribution over

- $\mathbb{G}_{n,d}$: d-regular graphs
- $\mathbb{CG}_{n,d}$: connected d-regular graphs
- $\mathbb{H}_{n,d}$: d-regular graphs with a complete Hamiltonian decomposition
- $\mathbb{I}_{n,d}$: d-regular graphs with a 1-factorization

Known Markov chains $\{G(t)\}_{t\in\mathbb{N}}$ with such distributions.

Examples of "Almost"-Uniform Distributions

Wormald [1999]: Uniform distribution over

- $\mathbb{G}_{n,d}$: *d*-regular graphs
- $\mathbb{CG}_{n,d}$: connected *d*-regular graphs
- $\mathbb{H}_{n,d}$: *d*-regular graphs with a complete Hamiltonian decomposition
- $\mathbb{I}_{n,d}$: *d*-regular graphs with a 1-factorization

Known Markov chains $\{G(t)\}_{t\in\mathbb{N}}$ with such distributions.
For $d \geq 3$, all of the above are expanders *a.a.s.*.

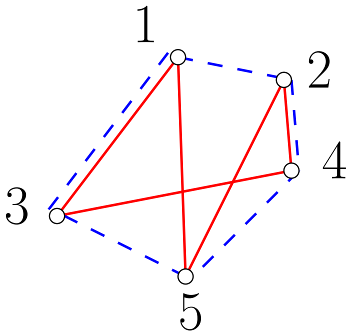
Examples of “Almost”-Uniform Distributions

Wormald [1999]: Uniform distribution over

- $\mathbb{G}_{n,d}$: d-regular graphs
- $\mathbb{CG}_{n,d}$: connected d-regular graphs
- $\mathbb{H}_{n,d}$: d-regular graphs with a complete Hamiltonian decomposition
- $\mathbb{I}_{n,d}$: d-regular graphs with a 1-factorization

Known Markov chains $\{G(t)\}_{t\in\mathbb{N}}$ with such distributions.
For $d \geq 3$, all of the above are expanders *a.a.s.*.
For $d > 5$, the latter two are expanders *w.h.p.*

Construction of a $2d$-regular expander in $\mathbb{MH}_{n,d}$ [Law and Siu, 2003]

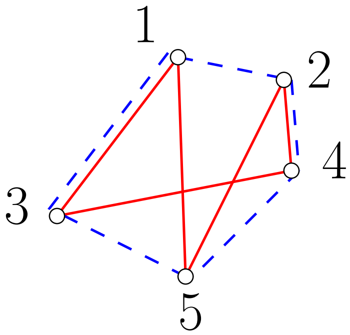


$$c_1 = [1, 3, 4, 2, 5]$$

$$c_2 = [5, 3, 1, 2, 4]$$

The multi-graph consists of d superimposed Hamiltonian cycles c_i.

Construction of a $2d$-regular expander in $\mathbb{MH}_{n,d}$ [Law and Siu, 2003]

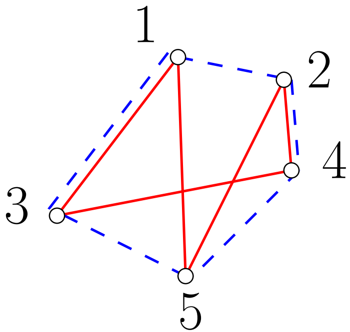


$c_1 = [1, 3, 4, 2, 5]$

$c_2 = [5, 3, 1, 2, 4]$

Each node has degree $2d$.

Construction of a 2d-regular expander in $\mathbb{MH}_{n,d}$ [Law and Siu, 2003]

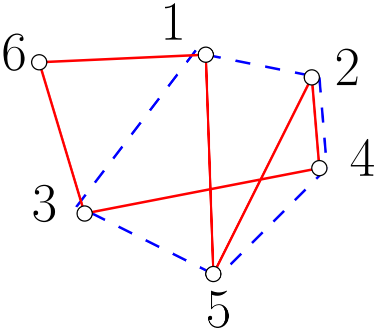


$c_1 = [1, 3, 4, 2, 5]$

$c_2 = [5, 3, 1, 2, 4]$

Each peer knows only its neighbors (d successors and d predecessors)

Construction of a $2d$-regular expander in $\mathbb{MH}_{n,d}$ [Law and Siu, 2003]

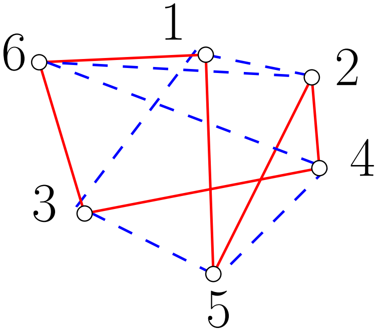


$c_1 = [1, 6, 3, 4, 2, 5]$

$c_2 = [5, 3, 1, 2, 4]$

For each cycle c_i, an incoming peer chooses a random peer and becomes its successor in c_i.

Construction of a $2d$-regular expander in $\mathbb{MH}_{n,d}$ [Law and Siu, 2003]

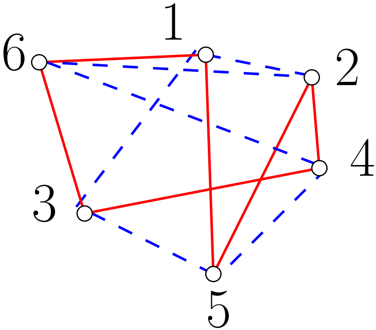


$c_1 = [1, 6, 3, 4, 2, 5]$

$c_2 = [5, 3, 1, 2, 6, 4]$

For each cycle c_i, an incoming peer chooses a random peer and becomes its successor in c_i.

Construction of a $2d$-regular expander in $\mathbb{MH}_{n,d}$ [Law and Siu, 2003]

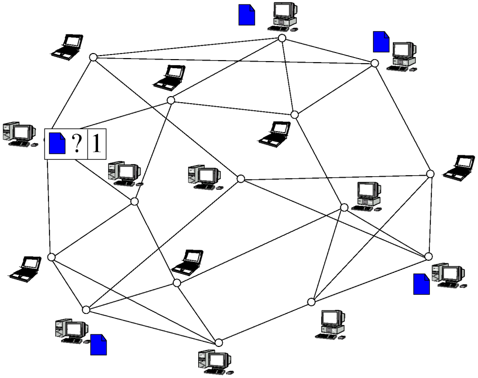


$$c_1 = [1, 6, 3, 4, 2, 5]$$

$$c_2 = [5, 3, 1, 2, 6, 4]$$

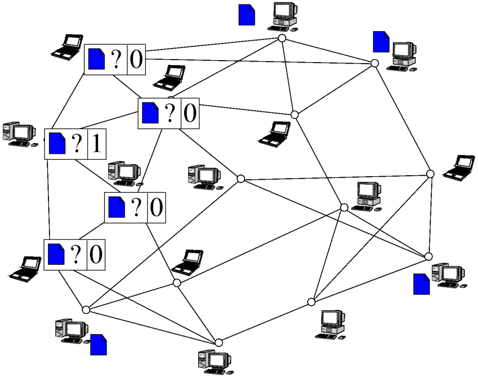
For $d > 5$ The resulting graph is an expander *w.h.p.*

Expanding Ring



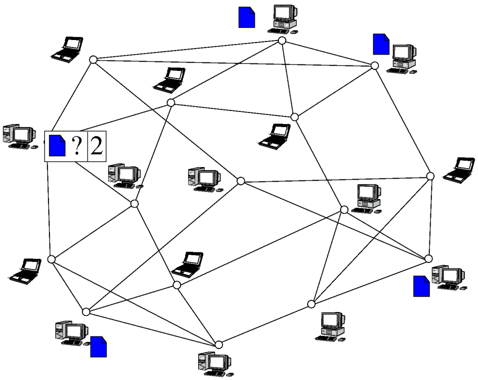
Query header initialized to 1

Expanding Ring



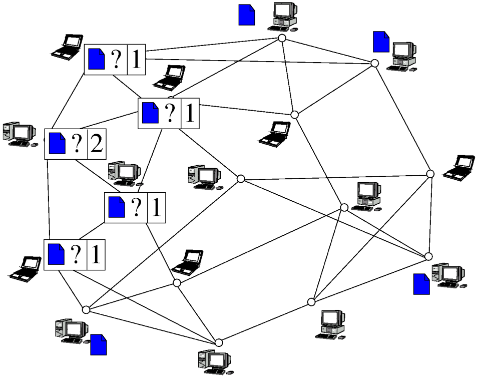
Query forwarded to all neighbours until it expires

Expanding Ring



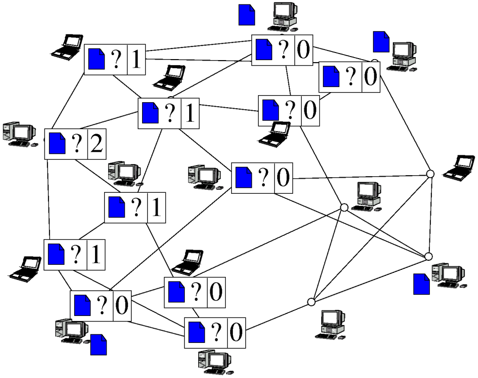
Process repeated with higher header value until either file located or initial value exceeds TTL_n

Expanding Ring



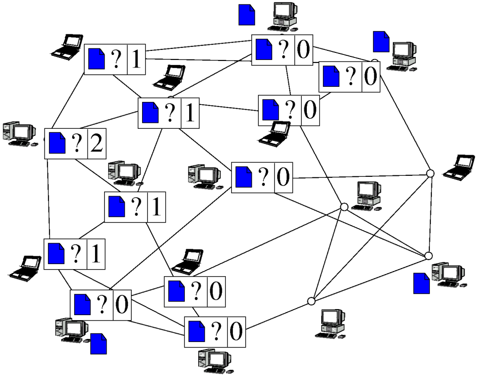
Process repeated with higher header value until either file located or initial value exceeds TTL_n

Expanding Ring



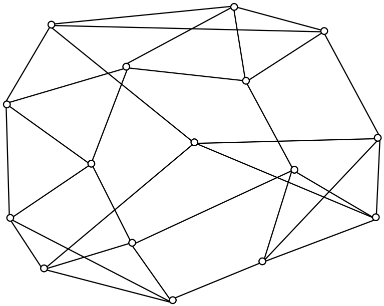
Process repeated with higher header value until either file located or initial value exceeds TTL_n

Expanding Ring



For both random walk and expanding ring, TTL_n is the maximum possible hop radius

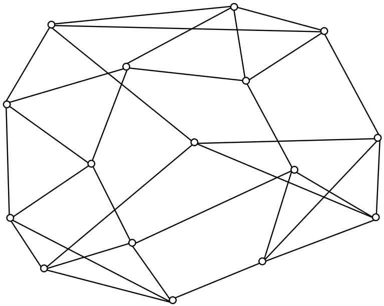
Overlay Graph: Churn-Driven Markov Chain



$\{G(t)\}_{t \in \mathbb{N}}$

$G(t)$: overlay graph at t-th departure/arrival epoch.

Overlay Graph: Churn-Driven Markov Chain



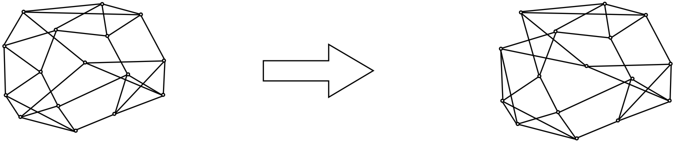
$$\{G(t)\}_{t\in\mathbb{N}}$$

$$G(t) \in \mathbb{G}_{n,d}$$

For all $t \geq 0, G(t)$ is d-regular graph with n vertices.

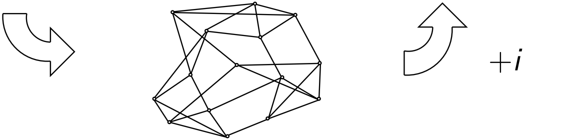
Overlay Graph: Churn-Driven Markov Chain

$G(t) = G$ $\qquad$ $G(t+1) = G'$



$$p^i_{GG'} = \mathbf{P}(G(t+1) = G' \mid G(t) = G, i)$$

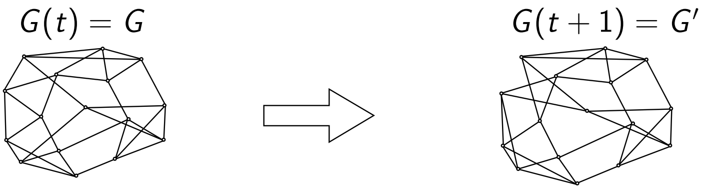
$-i$



$+i$

Transition from $G(t)$ to $G(t+1)$ depends on which peer is being replaced

Overlay Graph: Churn-Driven Markov Chain



$$p^i_{GG'} = \mathbf{P}(G(t+1) = G' \mid G(t) = G, i)$$

$$p_{GG'} = \mathbf{P}(G(t+1) = G' \mid G(t) = G) = \frac{1}{n}\sum_{i=1}^{n} p^i_{GG'}$$

$\{G(t)\}_{t\in\mathbb{N}}$ is a Markov chain with state space $\mathbb{S}_{n,d} \subseteq \mathbb{G}_{n,d}$.

Goal

Purpose:

- alleviate traffic load on server . . .
- . . . without overwhelming peers (clients).

Goal

Purpose:

- alleviate traffic load on server ...
- ...without overwhelming peers (clients).

Question: Is it possible to bound both

- the average traffic load per peer ρ_n?
- the server traffic load ρ_n^0?

Hybrid System: Random Walk with $\mathrm{TTL}_n = \Theta(n)$

Theorem

Assume that a graph sampled from the stationary distribution of $\{G(t)\}_{t\in\mathbb{N}}$ is an expander w.h.p.. Then

$$\rho_n = O(1) \quad \textit{and} \quad \rho_n^0 = O(1),$$

i.e., both loads generated by a random walk with $\mathrm{TTL}_n = \Theta(n)$ are bounded in n, irrespectively of p_n,q_n.

Hybrid System: Random Walk with $\mathrm{TTL}_n = \Theta(n)$

Theorem

Assume that a graph sampled from the stationary distribution of $\{G(t)\}_{t\in\mathbb{N}}$ is an expander w.h.p.. Then

$$\rho_n = O(1) \quad \textit{and} \quad \rho_n^0 = O(1),$$

i.e., both loads generated by a random walk with $\mathrm{TTL}_n = \Theta(n)$ are bounded in n, irrespectively of p_n,q_n.

- The hybrid system alleviates load at the server without overwhelming the peers!

Hybrid System: Random Walk with $\mathrm{TTL}_n = \Theta(n)$

Theorem

Assume that a graph sampled from the stationary distribution of $\{G(t)\}_{t\in\mathbb{N}}$ is an expander w.h.p.. Then

$$\rho_n = O(1) \quad \textit{and} \quad \rho_n^0 = O(1),$$

i.e., both loads generated by a random walk with $\mathrm{TTL}_n = \Theta(n)$ are bounded in n, irrespectively of p_n,q_n.

- The hybrid system alleviates load at the server without overwhelming the peers!
- Proved using results by Aldous and Fill

Hybrid System: Expanding Ring with $\mathrm{TTL}_n = \Theta(\log n)$

Worst-case response time is linear in n

Theorem

Assume that the stationary distribution of $\{G(t)\}_{t \in \mathbb{N}}$ is contiguous to the uniform distribution over $\mathbb{G}_{n,d}$. Then, there exists a $\mathrm{TTL}_n = \Theta\left(\log_{(d-1)} n\right)$ such that the expanding ring has

$$\rho_n = O\left(n^{\frac{\log(d-1)}{\log(d-3)}-1}\right) \text{ and } \rho_n^0 = O\left(n^{1-\frac{\log(d-3)}{\log(d-1)}}\right)$$

irrespectively of p_n, q_n.

Hybrid System: Expanding Ring with $\mathrm{TTL}_n = \Theta(\log n)$

Worst-case response time is linear in n

Theorem

Assume that the stationary distribution of $\{G(t)\}_{t\in\mathbb{N}}$ is contiguous to the uniform distribution over $\mathbb{G}_{n,d}$. Then, there exists a $\mathrm{TTL}_n = \Theta\left(\log_{(d-1)} n\right)$ such that the expanding ring has

$$\rho_n = O\left(n^{\frac{\log(d-1)}{\log(d-3)}-1}\right) \text{ and } \rho_n^0 = O\left(n^{1-\frac{\log(d-3)}{\log(d-1)}}\right)$$

irrespectively of p_n, q_n.

- Load growth is very slow —$O(n^{0.0199})$ for $d = 32$.
- Worst-case response time is $O\left(\log^2 n\right)$.

Hybrid System: Expanding Ring with $\mathrm{TTL}_n = \Theta(\log n)$

Worst-case response time is linear in n

Theorem

Assume that the stationary distribution of $\{G(t)\}_{t\in\mathbb{N}}$ is contiguous to the uniform distribution over $\mathbb{G}_{n,d}$. Then, there exists a $\mathrm{TTL}_n = \Theta\left(\log_{(d-1)} n\right)$ such that the expanding ring has

$$\rho_n = O\left(n^{\frac{\log(d-1)}{\log(d-3)}-1}\right) \text{ and } \rho_n^0 = O\left(n^{1-\frac{\log(d-3)}{\log(d-1)}}\right)$$

irrespectively of p_n, q_n.

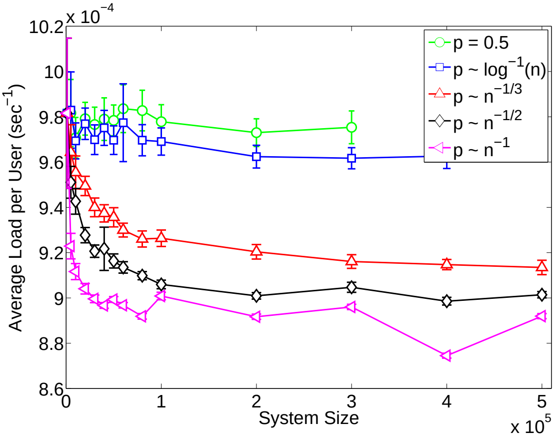
- Load growth is very slow —$O(n^{0.0199})$ for $d = 32$.
- Worst-case response time is $O\left(\log^2 n\right)$.
- Proved using results by Hoory et al. [2006].

Simulation Setup

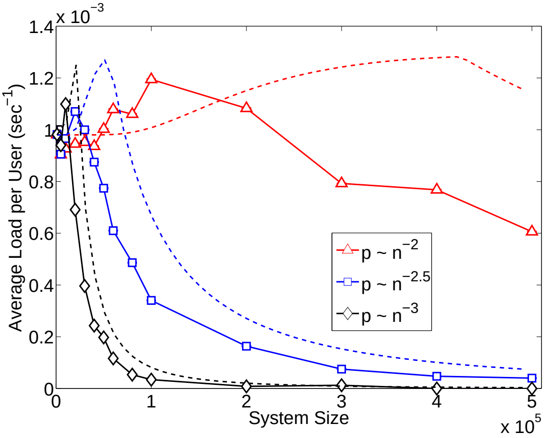
Simulations of Law and Siu [2003] peer-to-peer system:

- $\frac{1}{\mu} = 20\text{min}$.
- Arrival rate $n \cdot \mu$, $n = 10$ thousand to half a million.
- $\delta = 20\text{msec}$.
- $\mathrm{TTL}_n = n\delta$.
- Degree 16.

Load per peer ρ for popular items $(p = \omega(1/n))$.



Load per peer ρ for unpopular items ($p = O(1/n)$).



Server load ρ_0 for popular items $(p = \omega(1/n))$

Server load ρ_0 for popular items $(p = \omega(1/n))$

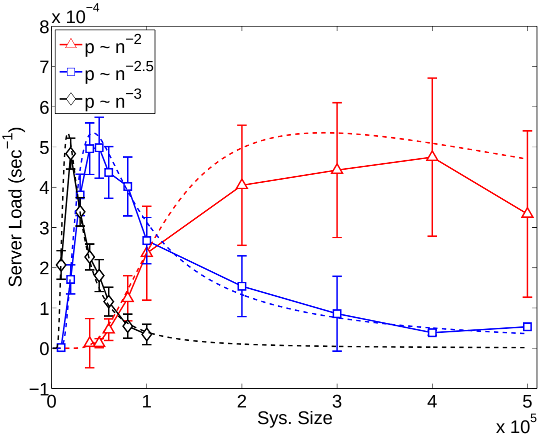
We saw none!!!!

Server load ρ_0 for popular items ($p = \omega(1/n)$)

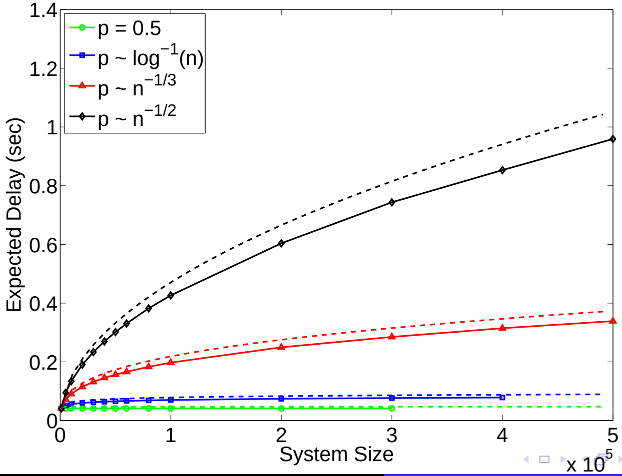
We saw none!!!!

Theoretical bound: $\rho_0 \sim 10^{-120}$

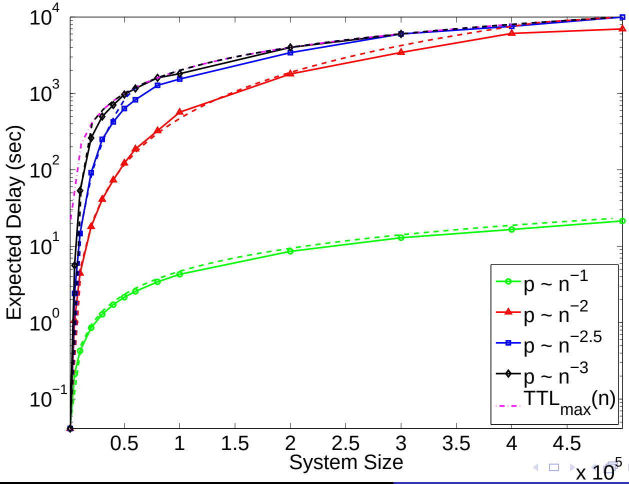
Server load ρ_0 for unpopular items ($p = O(1/n)$).



Delay for popular items ($p = \omega(1/n)$).



Delay for unpopular items ($p = o(1/n)$).



References I

Gnutella Measurement Studies

William Acosta and Surendar Chandra. Understanding the practical limits of the Gnutella p2p system: An analysis of query terms and object name distributions. In *MMCN*, 2008.

Chunxi Li and Changjia Chen. On Gnutella topology dynamics by studying leaf and ultra connection jointly in phase space. *Computer Networks*, 52(3):695–719, 2008.

Amir H. Rasti, Daniel Stutzbach, and Reza Rejaie. On the long-term evolution of the two-tier Gnutella overlay. In *INFOCOM*, 2006.

Matei Ripeanu, Adriana Iamnitchi, and Ian Foster. Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing*, 6(1), 2002.

Stefan Saroiu, Krishna P. Gummadi, and Steven D. Gribble. A measurement study of peer-to-peer file sharing systems. In *MMCN*, 2002.

Daniel Stutzbach, Reza Rejaie, and Subhabrata Sen. Characterizing unstructured overlay topologies in modern p2p file-sharing systems. *IEEE/ACM Transactions on Networking*, 16(2), 2008.

References II

Search Mechanisms

Yatin Chawathe, Sylvia Ratnasamy, Lee Breslau, Nick Lanham, and Scott Shenker. Making Gnutella-like p2p systems scalable. In *SIGCOMM*, 2003.

Edith Cohen and Scott Shenker. Replication strategies in unstructured peer-to-peer networks. *SIGCOMM Comput. Commun. Rev.*, 32(4):177–190, 2002. ISSN 0146-4833.

Christos Gkantsidis, Milena Mihail, and Amin Saberi. Hybrid search schemes for unstructured peer-to-peer networks. In *INFOCOM*, 2005.

Qin Lv, Pei Cao, Edith Cohen, Kai Li, and Scott Shenker. Search and replication in unstructured peer-to-peer networks. In *ICS*, 2002.

Krishna P.N. Puttaswamy, Alessandra Sala, and Ben Y. Zhao. Searching for rare objects using index replication. In *INFOCOM*, 2008.

Wesley W. Terpstra, Jussi Kangasharju, Christof Leng, and Alejandro P. Buchman. BubbleStorm: Resilient, probabilistic and exhaustive peer-to-peer search. In *SIGCOMM*, 2007.

Saurabh Tewari and Leonard Kleinrock. Proportional replication in peer-to-peer networks. In *INFOCOM*, 2006.

References III

Expander Graphs

David Aldous and Jim Fill. Reversible Markov Chains and Random Walks on Graphs. Monograph in preparation. http://www.stat.berkeley.edu/~aldous/RWG/book.html. Accessed on 29/12/2008.

Fan Rong K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

Joel Friedman. A proof of Alon's second eigenvalue conjecture. In *STOC '03*, pages 720–724, New York, NY, USA, 2003. ACM Press.

Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the AMS*, 43 (4):439–561, October 2006.

Markovian Overlay Graph Models

Tomas Feder, Adam Guetz, Milena Mihail, and Amin Saberi. A local switch markov chain on given degree graphs with application in connectivity of peer-to-peer networks. In *FOCS*, pages 69–76, 2006.

Ayalvadi J. Ganesh, Anne-Marrie Kermarrec, Erwan Le Merrer, and Laurent Massoulié. Peer counting and sampling in overlay networks based on random walks. *Journal of Distributed Computing*, 20(4), 2007.

Christos Gkantsidis, Milena Mihail, and Amin Saberi. Random walks in peer-to-peer networks. In *INFOCOM*, 2004.

Ching Law and Kai-Yeung Siu. Distributed construction of random expander networks. In *INFOCOM*, 2003.

Peter Mahlmann and Christian Schindelhauer. Peer-to-peer networks based on random transformations of connected regular undirected graphs. In *SPAA '05: Proceedings of the seventeenth annual ACM symposium on Parallelism in algorithms and architectures*, pages 155–164. ACM, 2005.

References IV

Replication

Edith Cohen and Scott Shenker. Replication strategies in unstructured peer-to-peer networks. *SIGCOMM Comput. Commun. Rev.*, 32(4):177–190, 2002. ISSN 0146-4833.

Qin Lv, Pei Cao, Edith Cohen, Kai Li, and Scott Shenker. Search and replication in unstructured peer-to-peer networks. In *ICS*, 2002.

Saurabh Tewari and Leonard Kleinrock. Proportional replication in peer-to-peer networks. In *INFOCOM*, 2006.

Miscellaneous

Michel Benaïm and Jean-Yves Le Boudec. A class of mean field interaction models for computer and communication systems. Technical Report LCA-REPORT-2008-010, EPFL, 2008.

Nicholas C. Wormald. Models of random regular graphs. *Surveys in Combinatorics*, 276:239–298, 1999.