Network Capacity

Lemma 1. *Let* $C > 3D$*; then at any time* $t \geq a \log N$*, (for some fixed constant* $a > 0$*), w.h.p. there are*

$$(1 - \frac{2D}{C-D}) \min[t, N](1 - o(1))$$

d-nodes in the network.

Proof. Let $t \geq N$.

Consider the interval $[t - N, t]$.

W.h.p. the number of new d-nodes arriving during this interval is $N(1 \pm o(1))$, and the number of connections to cache nodes from the new arrivals is $DN(1 \pm o(1))$.

The expected number of connections to the cache nodes in unit time in this interval is bounded by

$$\sum_{v \in V} (1 + o(1)) \frac{d(v)}{N} \frac{D}{d(v)} = D(1 + o(1)).$$

W.h.p. number of connections to the cache from old nodes in this interval is bounded by $ND(1+o(1))$.

Since a node receives $C - D$ connections while in the cache, w.h.p. no more than $\frac{2D}{C-D}N(1+o(1))$ d-nodes convert to new c-nodes in the interval.

Thus w.h.p we are left with $(1-\frac{2D}{C-D})N(1-o(1))$ d-nodes that joined the network in this interval. $\square$

Lemma 2. *Suppose that the cache is occupied at time t by node v. Let $Z(v)$ be the set of nodes that occupied the cache (in v's slot) during the interval $[t - c \log N, t]$.*

For any $\delta > 0$ and sufficiently large constant c, w.h.p. $|Z(v)|$ is in the range $\frac{2Dc}{(C-D)K} \log N(1 \pm \delta)$

Proof. The expected number of connections to a given cache node in an interval $[t - c \log N, t]$ is $\frac{2Dc \log N}{K}(1 + o(1))$.

W.h.p. the number of connections is in the range $\frac{2Dc}{K} \log N(1 \pm \delta)$.

Since a cache node receives $C - D$ connections while in the cache the result follows. $\square$

Efficient Cache Replacement

Lemma 3. *Assume that* $C > 3D$. *At any time* $t \geq c \log N$, *with probability* $1 - O(\frac{\log^2 N}{N})$ *the algorithm finds a replacement d-node by examining only* $O(\log N)$ *nodes.*

Proof. Let $v_1, ..., v_K$ be the K nodes in the cache at time t.

With probability $Ke^{-\frac{a \log^2 N}{N}} \geq 1 - O(\frac{\log^2 N}{N})$ no node in $Z(v_i)$, $i = 1, .., K$ leaves the network in the interval $[t - c \log N, t]$.

W.h.p. $Z(v)$ received at least $\frac{D}{K} c \log N$ connections from new d-nodes in the interval $[t - c \log N, t]$.

Among these new d-nodes no more than $|Z(v)|$ nodes entered the cache and became c-nodes during this interval.

Thus, w.h.p. there is a d-node attached to a node of $Z(v)$ at time t. $\square$

Connectivity

The proof that at any given time the network is connected w.h.p. is based on two properties of the protocol:

(1) Steps 3-4 of the protocol guarantee (deterministically) that at any given time a node is connected through "preferred connections" to a cache node;

(2) The random choices of new connections guarantee that w.h.p. the $O(\log N)$ neighborhoods of any two cache nodes are connected to each other.

Lemma 4. *At all times, each node in the network is connected to some cache node directly or through a path in the network.*

Lemma 5. *Consider two cache nodes v and u at time $t \geq c \log N$, for some fixed constant $c > 0$. With probability $1 - O(\frac{\log^2 N}{N})$ there is a path in the network at time t connecting v and u.*

Proof.

Let $Z(v)$ be the set of nodes that occupied the cache (in v's slot) during the interval $[t - c \log N, t]$. W.h.p. $|Z(v)| = d \log N$, for some constant d.

The probability that no node in $Z(v)$ leaves the network during the interval $[t - c \log N, t]$ is

$$e^{-\frac{cd \log^2 N}{N}} \geq 1 - O(\frac{\log^2 N}{N}).$$

The probability that no new node that arrives during the interval $[t - c \log N, t]$ connects to nodes in both $Z(v)$ and $Z(u)$ is bounded by $(1 - D^2/K^2)^{c \log N} = O(1/N^{c'})$. □

Theorem 1. *There is a constant c such that at any given time $t > c \log N$,*

$$Pr(G_t \text{ is connected}) \geq 1 - O(\frac{\log^2 N}{N}).$$

Corollary 1. *There is a constant c such that if the network is disconnected at time t,*

$$Pr(G_{t+c\log N} \text{ is connected}) \geq 1 - O(\frac{\log^2 N}{N}).$$

That is, the network rapidly recovers from fragmentation.

Theorem 2. *At any given time t such that $t/N \to \infty$, if the graph is not connected then it has a connected component of size $N(1 - o(1))$.*