Probable Security of Networks

LI Angsheng

Institute of Software
Chinese Academy of Sciences

Joint work with Yicheng Pan, Wei Zhang
Fragrant Hill Meeting
6th, Oct 2013

Outline

Infection set

Definition

(Infection set) Let $G = (V, E)$ be a network. Suppose that for each node $v \in V$, there is a threshold $\phi(v)$ associated with it. For an initial set $S \subset V$, the *infection set* of S in G is defined recursively as follows:

(1) Each node $x \in S$ is called *infected*.

(2) A node $x \in V$ becomes infected, if it has not been infected yet, and $\phi(x)$ fraction of its neighbors have been infected.

We use $\inf_G(S)$ to denote the infection set of S in G.

Thresholds of cascading

Definition

(Random threshold) We say that a cascading failure model is *random*, if for each node v, $\phi(v)$ is defined randomly and uniformly, that is, $\phi(v) = r/d$, where d is the degree of v in G, and r is chosen randomly and uniformly from $\{1, 2, \cdots, d\}$.

Definition

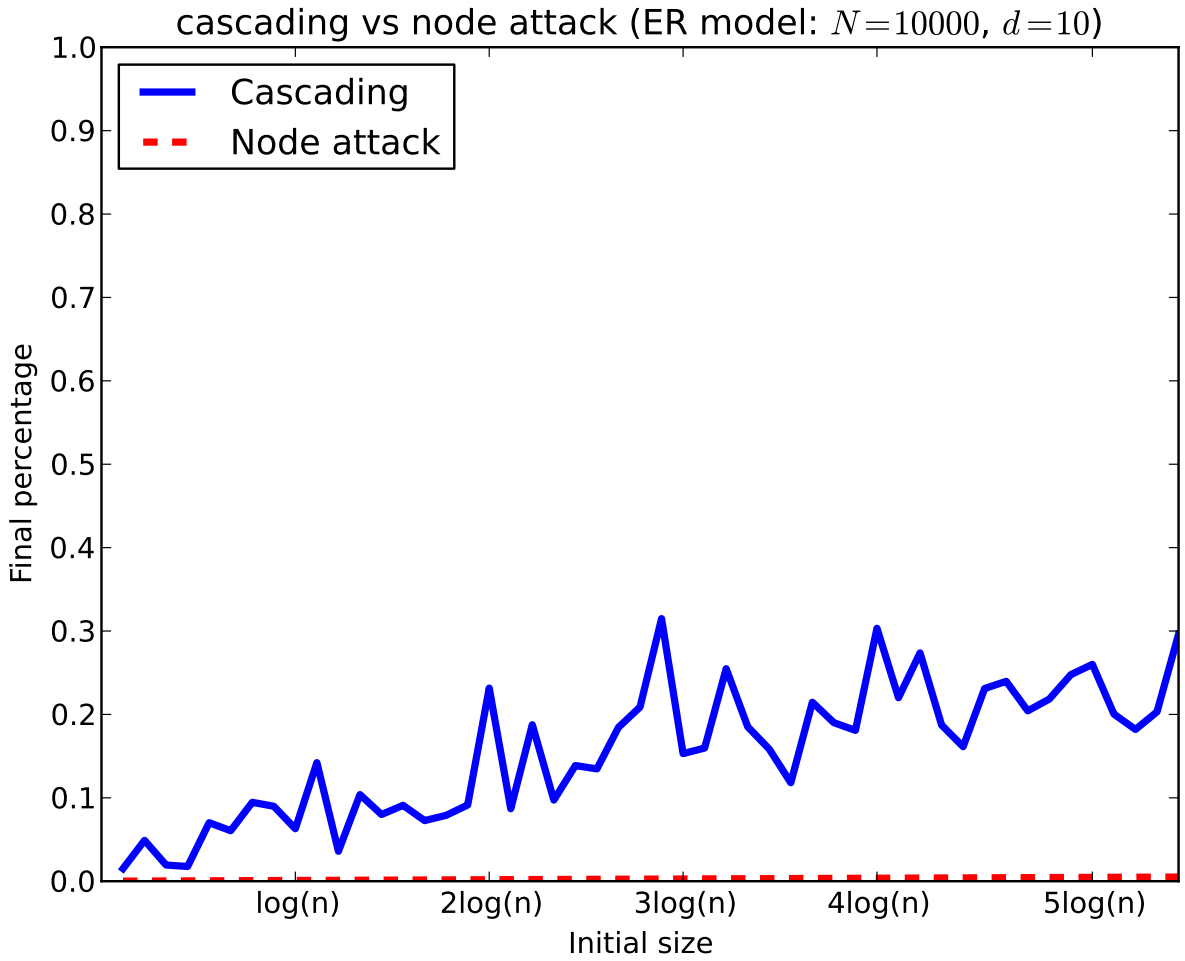
(Uniform threshold) We say that a cascading failure model is *uniform*, if for each node v, $\phi(v) = \phi$ for some fixed number ϕ.

Injury set

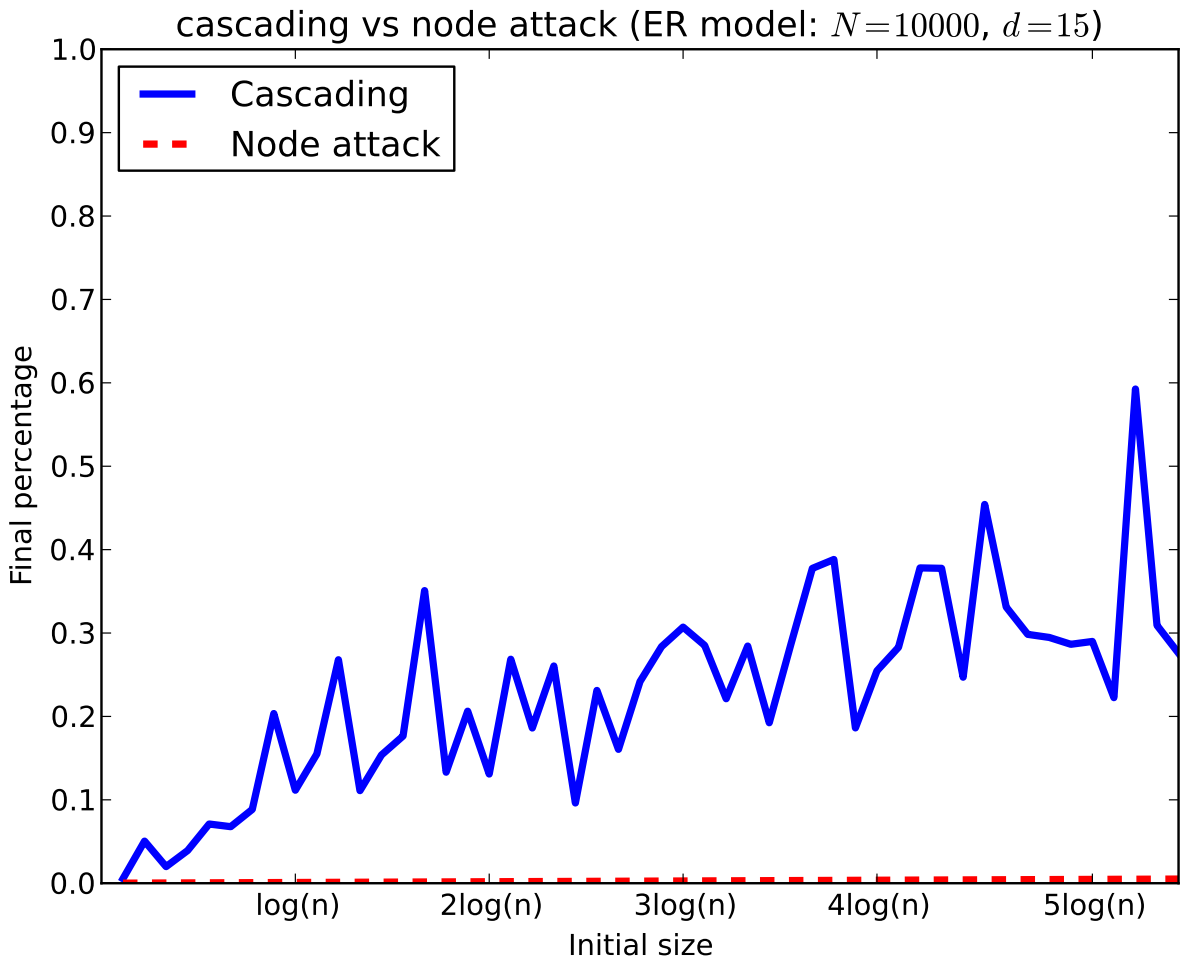
Definition

(Injury set) Let $G = (V, E)$ be a network, and S be a subset of V. The physical attacks on S is to delete all nodes in S from G. We say that a node v is injured by the physical attacks on S, if v is not connected to the largest connected component of the graph obtained from G by deleting all nodes in S.
We use $\mathrm{inj}_G(S)$ to denote the injury set of S in G.

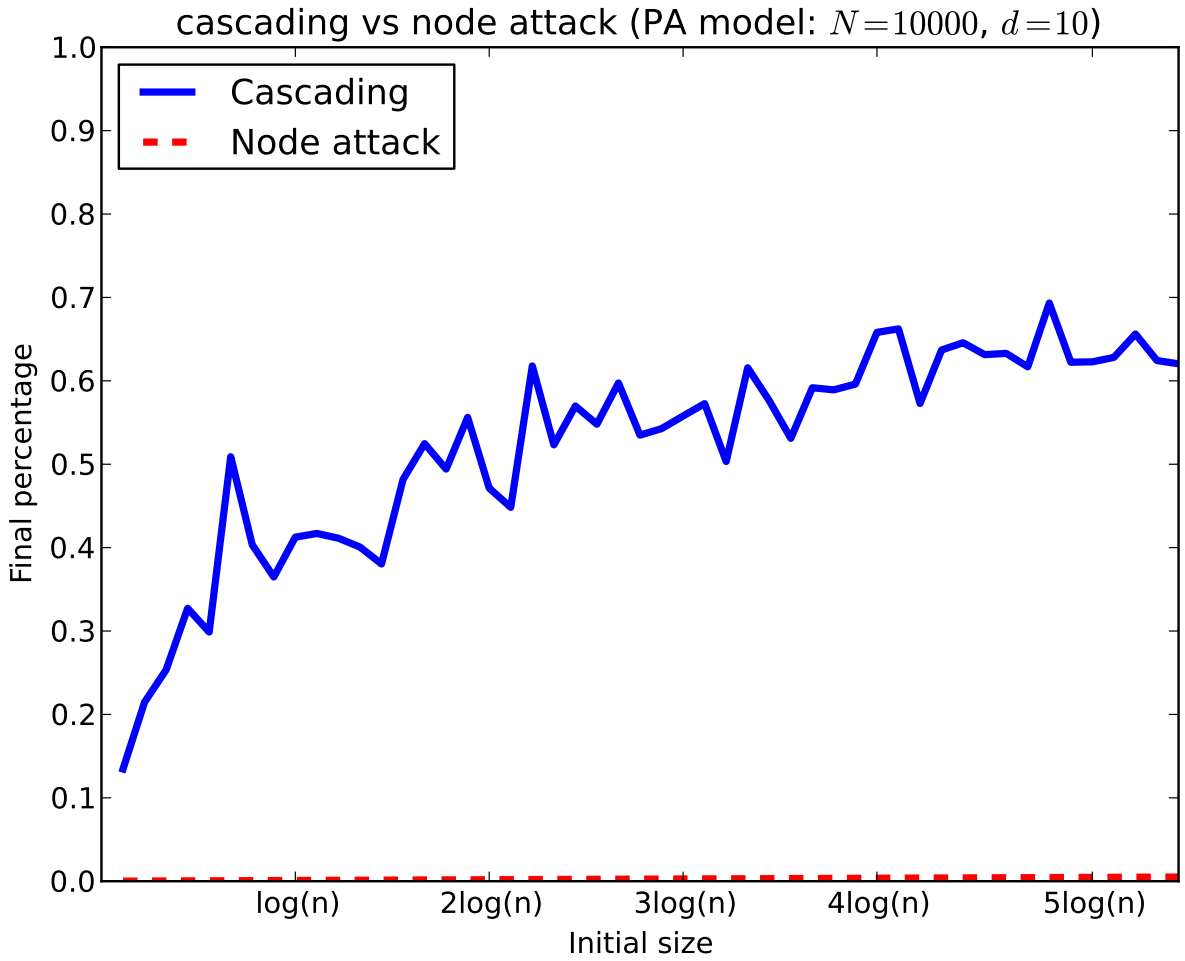
ER model


cascading vs node attack (ER model: $N=10000$, $d=10$)
Cascading
Node attack
Final percentage
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
log(n)
2log(n)
3log(n)
4log(n)
5log(n)
Initial size

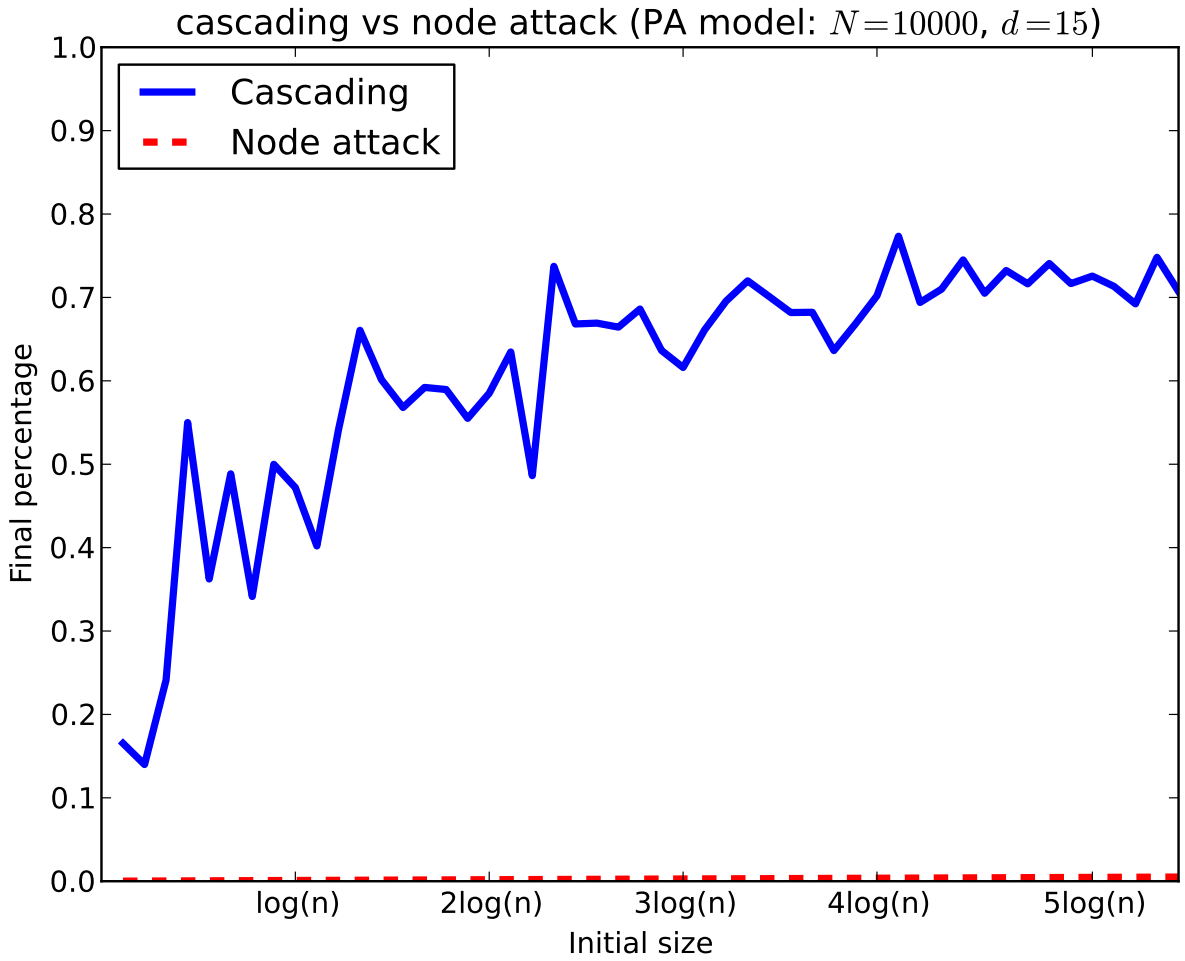
ER-2


cascading vs node attack (ER model: $N=10000$, $d=15$)
Cascading
Node attack
Final percentage
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
log(n)
2log(n)
3log(n)
4log(n)
5log(n)
Initial size

PA model


cascading vs node attack (PA model: $N=10000$, $d=10$)
Cascading
Node attack
Final percentage
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
log(n)
2log(n)
3log(n)
4log(n)
5log(n)
Initial size

PA-2


cascading vs node attack (PA model: $N=10000$, $d=15$)
Cascading
Node attack
Final percentage
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
log(n)
2log(n)
3log(n)
4log(n)
5log(n)
Initial size

Hypothesis

1. The infection sets are much larger than the corresponding injury sets.
   This means that to build our theory, we only need to consider the attacks of cascading failure models.
2. The attacks of top degree nodes of size as small as $O(\log n)$ may cause a constant fraction of nodes of the network to be infected under the cascading failure models of attacks.
   This means that networks of the ER and PA models are insecure for attacks of sizes as small as $O(\log n)$.

Random threshold security

Definition

(Random threshold security) For the cascading failure model of random threshold, we say that G is *secure*, if almost surely, meaning that with probability $1 - o(1)$, the following holds: for any set S of size bounded by a polynomial of $\log n$, the size of the infection set (or cascading failure set) of S in G is $o(n)$.

Uniform threshold security

Definition

(Uniform threshold security) For the cascading failure model of uniform threshold, we say that G is *secure*, if almost surely, the following holds: for an arbitrarily small ϕ, i.e., $\phi = o(1)$, for any set S of size bounded by a polynomial of $\log n$, S will not cause a global ϕ-cascading failure, that is, the size of the infection set of S in G, written by $\inf_G^{\phi}(S)$, is bounded by $o(n)$.

Questions

1. Can networks be secure?
2. What are the mechanisms of secure networks?

Security model

Definition

(Security model) Let $d \geq 4$ be a natural number and a be a real number, which is called *homophyly exponent*. We construct a network by stages.

(1) Let G_2 be an initial graph such that each node is associated with a distinct *color*, and called *seed*.

(2) Let $i > 2$. Suppose that G_{i-1} has been defined. Define $p_i = (\log i)^{-a}$.

(3) With probability p_i, v chooses a new color, c say. In this case, do:

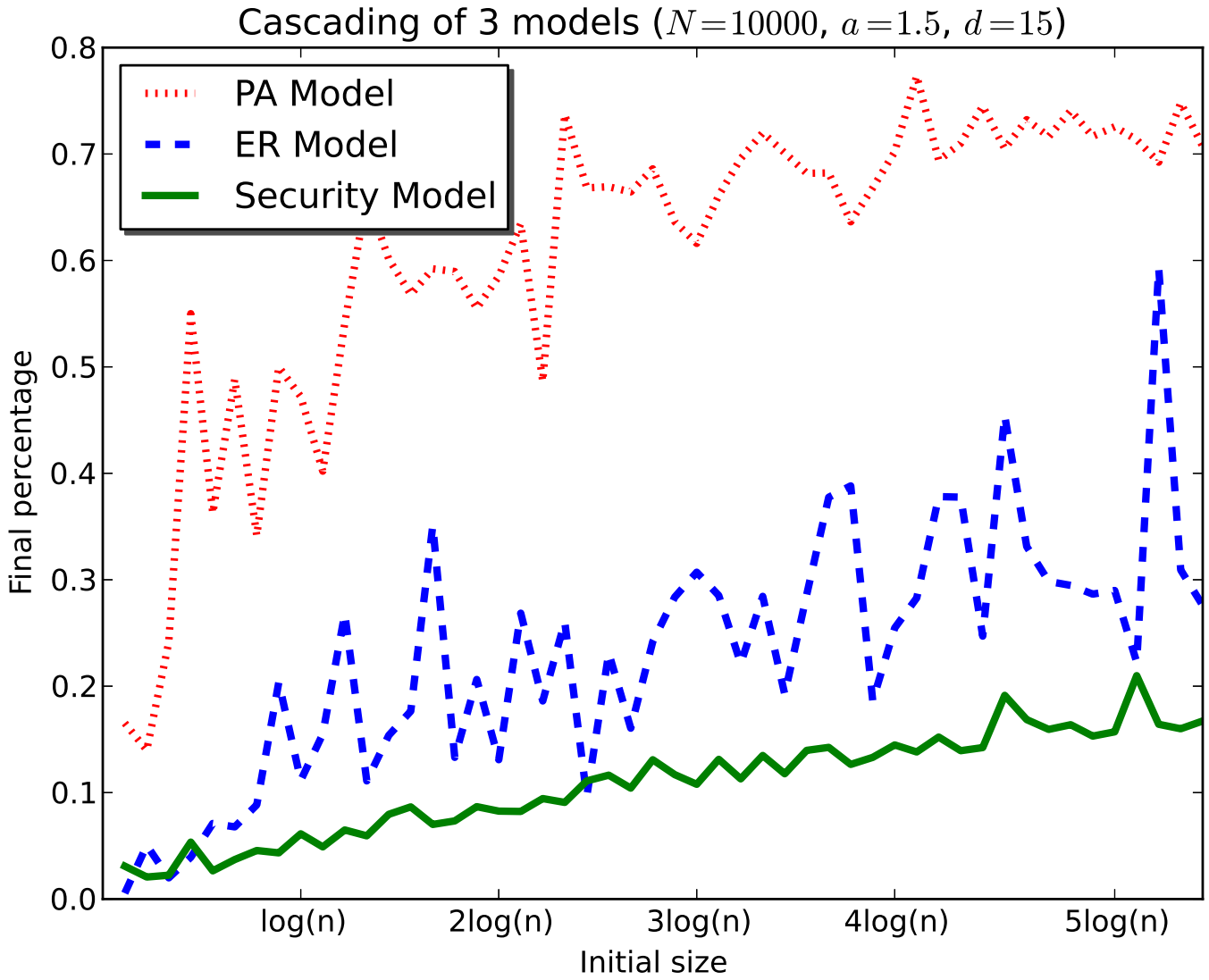
Security model-2

(3) 0.1 we say that v is the *seed node* of color c,
0.2 (Preferential attachment scheme) add an edge (u, v), such that u is chosen with probability proportional to the degrees of nodes in G_{i-1}, and
0.3 (Randomness) add $d-1$ edges (v, u_j), $j = 1, 2, \ldots, d-1$, where u_j's are chosen randomly and uniformly among all seed nodes in G_{i-1}.

(4) (Homophyly and preferential attachment) Otherwise. Then v chooses an old color, in which case, then:

0.1 let c be a color chosen randomly and uniformly among all colors in G_{i-1},
0.2 define the color of v to be c , and
0.3 add d edges (v, u_j), for $j = 1, 2, \ldots, d$, where u_j's are chosen with probability proportional to the degrees of all the nodes that have the same color as v in G_{i-1}.

Cascading in networks of security model



Networks of the security model are secure

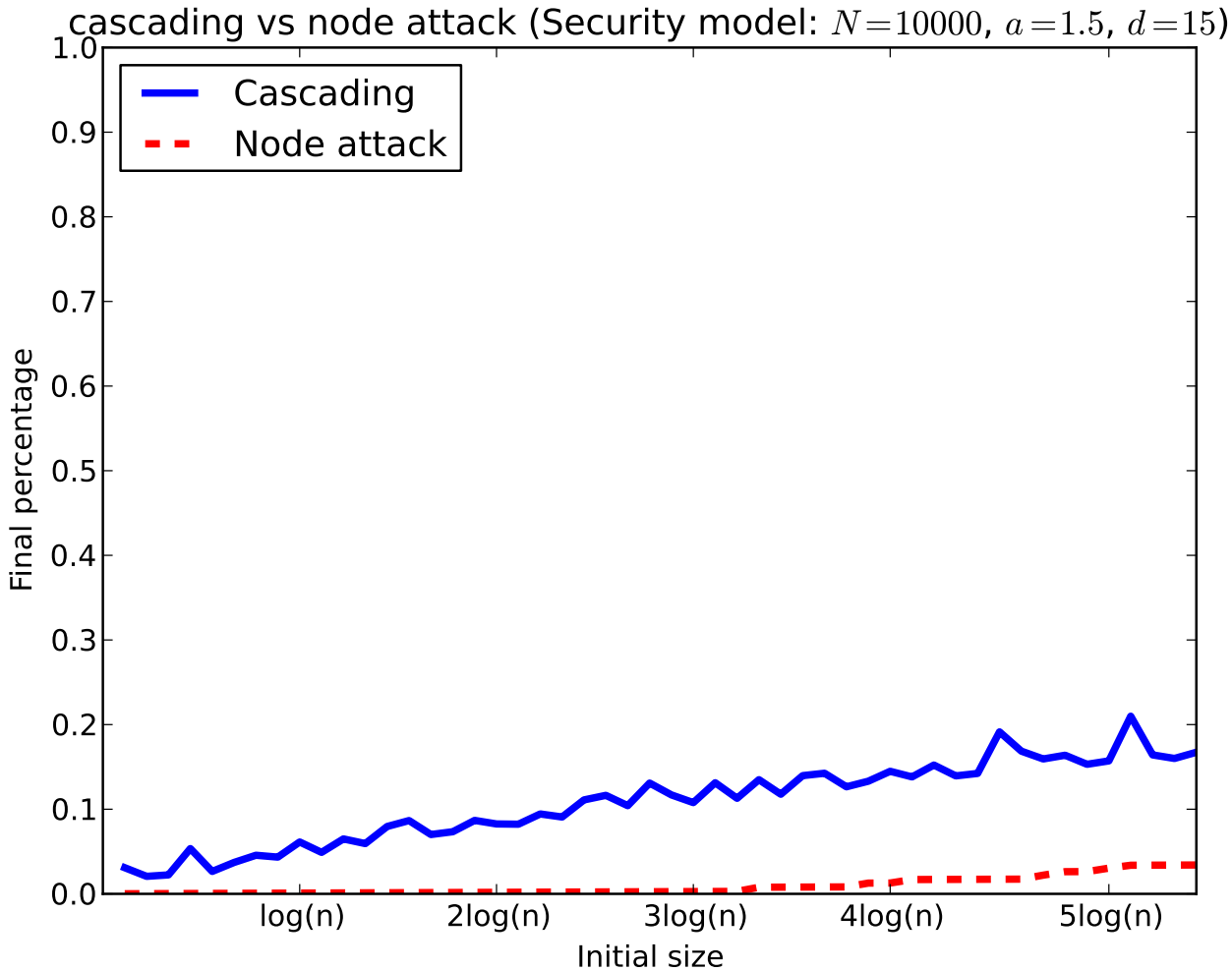


Figure: security model

Fundamental theorem

Theorem

(Fundamental principle) Let $a > 1$ *and* $d \geq 4$*. Then with probability* $1 - o(1)$*:*

(1) *(Basic properties):*

(i) *(Number of seed nodes is large) The number of seed nodes is bounded in the interval* $[\frac{n}{2\log^a n}, \frac{2n}{\log^a n}]$*.*

(ii) *(Communities whose vertices are interpretable by common features are small) Each homochromatic set has a size bounded by* $O(\log^{a+1} n)$*.*
We interpret a community by the common features of nodes in the community. This means that a community with interesting interpretations is small.

Fundamental - 2

(2) For degree distributions, we have:

(i) (Internal centrality) The degrees of the induced subgraph of a homochromatic set follow a power law.

(ii) The degrees of nodes of a homochromatic set follow a power law.

(iii) (Power law) Degrees of nodes in V follow a power law.

(iv) (Holographic law) The power exponents in (i) - (iii) above are the same.
This shows that the power exponent of a natural community is the same as that of the whole network.

Fundamental -3

(3) For node-to-node distances, we have:

(i) (Local communication law) The induced subgraph of a homochromatic set has a diameter bounded by $O(\log \log n)$.
This means that most communications in a network are local ones which are exponentially shorter than that of the global communications in the network.

(ii) (Small world phenomenon) The average node to node distance of G is bounded by $O(\log n)$.

Community structure principle

Theorem

For $a > 1$ and $d \geq 4$. Then with probability $1 - o(1)$:

(1) *(Small community phenomenon) There are $1 - o(1)$ fraction of nodes of G each of which belongs to a homochromatic set, W say, $\Phi(W)$, is bounded by $O\left(\frac{1}{|W|^{\beta}}\right)$ for $\beta = \frac{a-1}{4(a+1)}$.*

(2) *(Conductance community structure theorem) The conductance community structure ratio of G is at least $1 - o(1)$, that is, $\theta(G) = 1 - o(1)$.*

(3) *(Modularity community structure theorem) The modularity of G is $1 - o(1)$, that is, $\sigma(G) = 1 - o(1)$.*

(4) *(Entropy community structure theorem) The entropy community structure ratio of G is $1 - o(1)$, that is, $\tau(G) = 1 - o(1)$.*

Degree priority

Definition

Given a node v,
Define the length of degrees of v to be the number of colors of the neighbors of v, written $l(v)$
For j, define the j-th degree of v to be the j-th largest number of edges from v to its homochromatic neighbors, written $d_j(v)$.

Degree priority principle

Theorem

(Degree priority principle) Then with probability $1 - o(1)$*:*

(1) *(First degree property) The first degree of* v*,* $d_1(v)$ *is the number of edges from* v *to nodes of the same color as* v*.*

(2) *(Second degree property) The second degree of* v *is bounded by a constant, i.e.,* $d_2(v) \leq O(1)$

(3) *If* v *is a seed node, then the first degree of* v*,* $d_1(v)$ *is at least* $\Omega(\log^{\gamma} n)$ *for some constant* γ*.*

Degree priority principle -2

Let $G = (V, E)$ be a network of the security model. Then with probability $1 - o(1)$, the following properties hold: Let N be the number of seed nodes in G. For $l = N^{1-\theta}$ and $r = \frac{N}{\log^c N}$ for some constants θ and c.

(1) Let x be a seed created before time step l. Then the length of degrees of x in G is at least $\Omega(\log n)$.

(2) Let y be a seed created before time step r. Then the length of degrees of y in G is at least $\Omega(\log \log n)$

(3) Let z be a seed created after time step r. Then the length of degrees of z in G is at most $O(\log \log n)$.

(4) For a randomly chosen x, the length of degrees of x is $l(v) = O(\log \log n)$.

(5) For ant seed v, $l(v) \geq d - 1$.

Almost all communities are strong

A community X is strong, if its seed x_0 say cannot be infected by collection of all nodes fail to share the same color as v, unless some node in X has already been infected.
With prob $1 - o(1)$, almost all communities are strong.

Widths principle

Theorem

(Widths Principle) For $l = N^{1-\theta}$ and $r = \frac{N}{\log^c N}$ for some constants θ and c. We say that a community is created at time step t, if the seed node of the community is created at time step t.

(1) *Let X be a community created before time step l. Then the width of X in G is at least $\Omega(\log n)$.*

(2) *Let Y be a community created before time step r. Then the width of Y in G is at least $\Omega(\log \log n)$*

(3) *Let Z be a community created after time step r. Then the width of Z in G is at most $O(\log \log n)$.*

(4) *For a randomly chosen X, the width of X in G is $w^G(X) = O(\log \log n)$.*

Inclusion and infection principle

Theorem

(Inclusion and infection principle) Let $G = (V, E)$ be a security network. Then for following properties hold:

(1) *(Inclusion) For a non-seed node x in G, the width of x in G is $w^G(x) = 0$.*

(2) *(Widths of seed nodes) For every seed node x in G, the width of x is at most 1.*

Infection priority tree

Define T:

1. delete all edges created by seeds to seeds chosen randomly
2. merge each community into a single node

Infection of a strong community must be intrigued by an edge in the infection priority tree T.

Infection priority tree principle

Theorem

With prob $1 - o(1)$*, the infection priority tree has height* $O(\log n)$*.*

Uniform threshold security theorem

Theorem

(Uniform threshold security theorem) Let G be a graph constructed from $\mathcal{S}(n, a, d)$ with $p_i = \log^{-a} i$ for homophyly exponent $a > 4$ and for $d \geq 4$. Let the threshold parameter $\phi = O\left(\frac{1}{\log^b n}\right)$ for $b = \frac{a}{2} - 2 - \epsilon$ for arbitrarily small $\epsilon > 0$. Then with probability $1 - o(1)$, we have that for any constant $c > 0$,

$$\Pr_{G \in_R \mathcal{S}(n,a,d),\ G=(V,E)} \left[\forall S \subseteq V,\ |S| = \lceil \log^c n \rceil,\ |\inf_G^{\phi}(S)| = o(n)\right]$$

$$= 1 - o(1).$$

Random threshold security theorem

Theorem

(Random threshold security theorem) Let $a > 6$ be the homophyly exponent, and $d \geq 4$. Suppose that G is a graph generated from $\mathcal{S}(n, a, d)$.
Then with probability $1 - o(1)$ (over the construction of G), there is no initial set of poly-logarithmic size which causes a cascading failure set of non-negligible size. Formally, we have that for any constant $c > 0$,

$$\Pr_{G \in_R \mathcal{S}(n,a,d),\ G=(V,E)} \left[\forall S \subseteq V, |S| = \lceil \log^c n \rceil, |\inf_G^R(S)| = o(n) \right]$$

$$= 1 - o(1).$$

Proof sketch

Let G be a network of the security model, and S be a set of attacks such that $|S|$ is bounded by a polynomial of $\log n$.

1. Let k be the number of vulnerable communities
2. There are at most $|S| + k$ nodes which intrigue a cascading procedure among the strong communities
3. By the infection priority tree principle, there are at most $O((|S| + k) \log n)$ communities that are infected by the attacks on S
4. By the fundamental theorem, there are at most

$$O((|S| + k) \cdot \log n \cdot \log^{a+1} n)$$

many nodes that are infected. The later could be $o(n)$.

Mechanisms

1. Homophyly and randomness are the mechanisms of security of power law networks
2. Power law and small world property are not obstacles of security of networks
3. Network security can be mathematically guaranteed
4. Hypothesis: nature solves security by mechanisms - social, biological and physical understanding of the security model - open

New principles

There are new principles of the security model solving fundamental problems such as the prisoner's dilemma in power law networks - in progress

Open questions

1. To develop a security theory of networks, many fundamental questions - open
2. To push our theory to practical applications - many algorithmic, engineering and cryptographical issues - open

Thank You