Size Estimation and Change Detection in Anonymous Networks

Lucchese Riccardo   Savietto Michele

Corso di Laurea in Ingegneria dell'Automazione
Università degli Studi di Padova

27 Luglio 2012

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Summary

1. Problem Definition
2. Static Network - Consensus Algorithms
3. Change Detection
4. Simulations

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

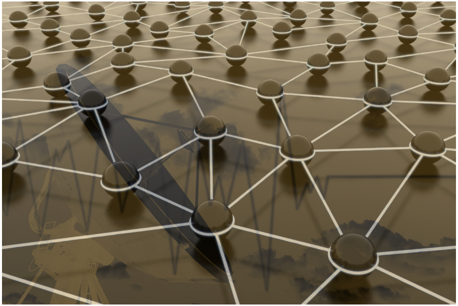
Problem Definition

Problem A:

Estimate the size S of an anonymous network exploiting consensus methods.

Problem B:

Formulate an hypothesis test for detecting topology changes.



DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Problem definition - A

The constraint of anonymity characterizes the problem in a critical way.

In fact, if one assumes no a-priori knowledge on the network topology, an impossibility results states that there is no estimation strategy with finite computational complexity that can output the correct size almost surely.

Proposed estimation strategies differ by how nodes generate statistical information on the network size. There are three main groups: random walks, sampling methods and consensus based methods. The estimation step reduces to Maximum Likelihood (ML) procedures.

Problem - B

Devise an hypothesis testing approach for detecting topology changes in a node neighborhood. The change-detection algorithm should be distributed such that nodes can run it locally.

We could not find any source on change detection in the scientific literature that dealt with the constraint of anonymity.

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Contribution

For the estimation strategy based on average consensus and common Bernoulli distribution we provided a refined bound on the estimator error probability that slightly improves the state of art. The results is, however, of little practical interest due too slow convergence rate and scalability of average consensus and to the high sensibility of the estimator statistic to noise superposed to the outcome of consensus.

We formulated a neighborhood size change detection test in the form of a Generalized Likelihood Ratio (GLR) and provided its statistical characterization in terms of type I and II error probabilities. We have shown that, if tuned opportunely, the test achieves good detection performances.

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Static Network

Fundamental hypothesis for this Section:

network topology have not to change, i.e. number of agents is constant.

Considered cases:

- Average Consensus, with Bernoulli distribution
- Max Consensus, with Uniform distribution

Network Size Estimation

Let us model a static connected network with a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where $\mathcal{N} \subset \mathbb{N}^+$ is the set of nodes and $\mathcal{E} \subset \mathcal{N}^2$ is the set of undirected edges such that $(i, j) \in \mathcal{E}$ if node i and node j can communicate.

The general estimation strategy in consensus based methods comprises three main steps.

i) each node $i \in \mathcal{N}$ starts by generating a local vector of initial random values, $\boldsymbol{y}^{(i)} \in \mathbb{R}^M$, by sampling M i.i.d. random variables with common probability density $p(\cdot)$.

ii) the network distributedly computes a consensus function of these initial values such that asymptotically each node reaches consensus on a quantity

$$\boldsymbol{f} = F\left(\left\{\boldsymbol{y}^{(i)} \ : \ i \in \mathcal{N}\right\}\right) ,$$

iii) each node statistically infers an estimate of $S = |\mathcal{N}|$ given $\boldsymbol{f}$.

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Average Consensus - Bernoulli distribution

Each agent i locally generates

$$\boldsymbol{y}^{(i)} \sim \mathcal{B}(p)$$

The consensus function is the average:

$$\boldsymbol{f} := \frac{1}{S} \sum_{i \in \mathcal{N}} \boldsymbol{y}^{(i)}$$

Since $Sf_m \sim \mathrm{Bin}(S, p)$, exploiting independency, we can fully determine the distribution of $\boldsymbol{f}$.

Suppose that f_m is observed. The likelihood of f_m as a function of S has (unbounded) support on the discrete set

$$\mathrm{supp}\,\ell\,(S\;;\;f_m) = \left\{\nu\bar{S}\;:\;\nu \in \mathbb{N}^+, f = \frac{\bar{k}}{\bar{S}}\ \text{with}\ (\bar{k},\bar{S})\ \text{coprime}\right\}$$

It can be shown that the ML estimator of S is

$$\widehat{S}(\boldsymbol{f}) = \min\left(\bigcap_{m=1}^{M}\mathrm{supp}\,\ell\,(S\;;\;f_m)\right)$$

$\widehat{S}(\boldsymbol{f})$ is thus the smallest size which can explain the M independent observations, $f_1, \ldots f_M$, in terms of $\boldsymbol{f} := \frac{1}{S}\sum_{i\in\mathcal{N}}\boldsymbol{y}^{(i)}$.

Difficult to provide a closed form for the estimator distribution because of the discrete nature of the initial distribution.

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Error Probability Characterization

Define $\alpha(p)$ the error probability. When $M = 1$ it can be bounded from above by

$$\overline{\alpha} = 1 - \phi(S)/S$$

where $\phi(\cdot)$ is the Euler phi-function.

Since $\phi(S)/S > 0.15$ at least up to $S = 10^{10}$, one can argue that as long as p is far from zero or one than $\alpha(p) < 0.85$.

An upper bound, for the general case $M \geq 1$, can be

$$\mathbb{P}\left[\widehat{S}(\boldsymbol{f}) \neq S \,;\ S\right] \leq \overline{\alpha}^M$$

Now we provide a refinement of this error probability.

The intuition behind our approach is that the estimator error-probability can be computed in closed form when S is a power of 2.

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Characterization of our bound - 1

Define $\psi : \mathbb{Q} \to \mathbb{N}^+$ as the application mapping each rational into the value taken by the denominator of its coprime representation (with convention $\psi(0) \mapsto 1$) and $d_m = \psi(f_m)$, $m = 1, \ldots, M$.

Consider the prime-number factorizations

$$S = 2^{\gamma_2} 3^{\gamma_3} 5^{\gamma_5} \ldots, \quad d_m = 2^{\gamma_2^m} 3^{\gamma_3^m} 5^{\gamma_5^m} \ldots \quad m = 1, \ldots, M$$

The previous ML estimator provides the correct size if and only if the lowest common multiple of the M observations $d_1, \ldots, d_M$ is exactly S.

This happens if and only if for each prime number $\nu \in \mathrm{primes}(S)$ there exist at least one $m \in \{1, \ldots, M\}$ such that ν^{γ_ν} divides d_m.

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Characterization of our bound - 2

Since the M variables $f_1, \dots, f_M$ are i.i.d., it can be proved that

$$\mathbb{P}\left[\widehat{S}(\boldsymbol{f}) \neq S \; ; \; S\right] \leq |\mathrm{primes}(S)| \, \beta(p)^M$$

where

$$\beta(p) := \frac{1 + (1 - 2p)^S}{2}$$

is the probability that the outcome of a binomial random variable with success probability p and a number of experiments S is even.

Remark

Our new bound is the exact error probability when S is a power of two.

$p = 0.5$ minimizes the worst case error probability over a certain range of possible sizes $[1, S_{max}]$. It corresponds to a global minimum of $\beta(p)$ for any S power of two.

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Comparison between old and new bounds

Our bound vanishes asymptotically faster in M than the one argumented in the existing articles.

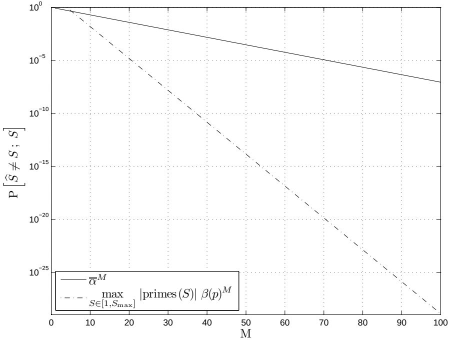


Figure: Parameters settings: $\overline{\alpha} = 0.85$; range of sizes $[1, 10^6]$ for which we have found numerically $\max_{S\in[1,10^6]} |\text{primes}(S)| = 16$; $p = 0.5$.

Max Consensus - Uniform distribution

Each agent i locally generates

$$\boldsymbol{y}^{(i)} \sim \mathcal{U}\left[0, 1\right]$$

The consensus function is the maximum

$$f_m := \max_{i \in \mathcal{N}} y_m^{(i)} \quad , \quad m = 1, \ldots, M$$

It can be shown that the joint probability density of $\boldsymbol{f}$ is

$$p\left(\boldsymbol{f}\ ;\ S\right) = \begin{cases} S \prod_{m=1}^{M} f_m^{S-1} & 0 \leq f_m \leq 1\ ,\ m = 1, \ldots, M \\ 0 & \text{otherwise} \end{cases}$$

DEPARTMENT OF INFORMATION ENGINEERING UNIVERSITY OF PADOVA

It follows that the ML estimator of S is

$$\widehat{S}(\boldsymbol{f}) = \arg\max_{S \in \mathbb{R}^+} p\left(\boldsymbol{f}\ ;\ S\right) = \frac{M}{\sum_{m=1}^{M} - \log f_m}$$

Except for a scaling factor, $\widehat{S}(\boldsymbol{f}) \sim \Gamma^{-1}$ and for $M > 2$, the estimator relative mean square error is only function of M. In particular it decreases when M becomes bigger.

The max-consensus estimation strategy is easily extended to network with dynamic topology.

Neighborhood size estimation

We briefly recall a generalization of the max-consensus estimator which is needed for our approach to change detection. The new estimation procedure assumes that the process by which the network changes topology is deterministic (and unknown).

Communication protocol: the time is divided in epochs, indexed by $t = 0, 1, 2, \ldots$. Every agent broadcasts its information exactly once per epoch (the order is irrelevant).

Network model: at each epoch t the we describe the network by means of a graph, $\mathcal{G}(t) := (\mathcal{V}(t), \mathcal{E}(t))$, where $\mathcal{V}(t)$ is the set of agents active at time t and $\mathcal{E}(t) \subseteq \mathcal{V}(t) \times \mathcal{V}(t)$ is the set of communications among active agents at time t: $(i, j) \in \mathcal{E}(t)$ indicates that agent i has successfully broadcast its information to j.

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Information generation scheme

for $t = 1, 2, \ldots$ **do**

(Information Update) each agent i computes $F^{(i)}(t)$ by shifting the columns of $F^{(i)}(t-1)$, in the sense that $\boldsymbol{f}_k^{(i)}(t) = \boldsymbol{f}_{k-1}^{(i)}(t-1)$ for $k = 2, \ldots, D$. $\boldsymbol{f}_1^{(i)}(t)$ is instead filled with M new i.i.d. random values $f_{m,1}^{(i)}(t) \sim \mathcal{U}[0,1]$, $m = 1, \ldots, M$

(Communication) every agent broadcasts $F^{(i)}(t)$ to its neighbors

(Information Mixing) each agent i updates its $F^{(i)}(t)$ by means of the $F^{(j)}(t)$'s received from its neighbors. More specifically

$$f_{m,k}^{(i)}(t) \leftarrow \max_{(j,i) \in \mathcal{E}(t)} \left(f_{m,k}^{(i)}(t), \left\{ f_{m,k}^{(j)}(t) \right\} \right)$$

for $m = 1, \ldots, M$, $k = 1, \ldots, D$.

end for

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

It is not difficult to see that the set of nodes from which $\boldsymbol{f}_k^{(i)}$ aggregates information is given for $k \geq 1$ by the recursion

$$\mathcal{V}_k^{(i)}(t) := \bigcup_{(j,i)\in\mathcal{E}(t)} \mathcal{V}_{k-1}^{(j)}(t-1)\,,$$

with initial conditions $\mathcal{V}_0^{(i)}(t) = \{i\}$.

It follows from the previous slides that the ML estimator of $S_k^{(i)}(t) = \left|\mathcal{V}_k^{(i)}(t)\right|$ is the statistic

$$\widehat{S_k^{(i)}}(t) := \frac{M}{\sum_{m=1}^{M} -\log\left(f_{m,k}^{(i)}(t)\right)}\,,$$

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Hypothesis testing

Let $\boldsymbol{f} \sim \{p_\theta\}_{\theta\in\Theta}$, where Θ is the the set of a-priori plausible values that θ might take, and consider two complementary hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ of the form

$$\mathcal{H}_i := \{\boldsymbol{f} \sim p_\theta \ , \ \theta \in \Theta_i\} \ , \quad \Theta_0 \cap \Theta_1 = \emptyset, \quad \Theta_0 \cup \Theta_1 = \Theta \ .$$

Deciding between $\mathcal{H}_0$ and $\mathcal{H}_1$ is performed by means of a deterministic decision rule g with range $\{0, 1\}$. Two different kinds of errors can occur:

- accepting $\mathcal{H}_1$ when $\mathcal{H}_0$ is true, a.k.a. type I error
- accepting $\mathcal{H}_0$ when $\mathcal{H}_1$ is true, a.k.a. type II error

It is meaningful to characterize g in terms of *type I and II error probabilities*.

Hypothesis testing

Given a decision function g, define the test *power function* as

$$\beta_g(\theta) = \mathbb{P}\left[\mathcal{H}_1 \text{ is accepted }; \ \theta\right] = \mathbb{E}\left[g(\boldsymbol{f}) \ ; \ \theta\right] \ .$$

Both type I and II error probabilities can be expressed by means of $\beta_g(\cdot)$. In particular, the *worst case* type I error probability, a.k.a the test *size*, is given by

$$\alpha_0 = \sup_{\theta \in \Theta_0} \beta_g(\theta) \ .$$

When the hypotheses are simple, it can be shown that in the class of decision functions with a given size α_0, the minimizers of the type II error probability take a special form.

Hypothesis testing

This special form is given by the Neymann-Person lemma

$$g(\boldsymbol{f}) := \begin{cases} 0 & \Lambda \leq \lambda \\ 1 & \text{otherwise} \end{cases}$$

where Λ is the Likelihood Ratio (LR)

$$\Lambda = \frac{\ell(\theta_0\ ;\ \boldsymbol{f})}{\ell(\theta_1\ ;\ \boldsymbol{f})}\ , \quad \theta_i \in \Theta_i\ ,$$

and λ is the *test threshold*: it realizes a trade-off between the test sensitivity to high frequency changes and the rate of false positives.

Change detection

Our aim is to detect disconnections in the k-steps neighborhoods over a temporal window spanning $N+1$ epochs. Intuitively this should be possible by inspecting the evolution of the samples

$$\boldsymbol{f}_k^{(i)}(t-N+1, t) = \left(\boldsymbol{f}_k^{(i)}(t-N), \ldots, \boldsymbol{f}_k^{(i)}(t)\right) \ .$$

We shall make the assumption that at most one sensible size change happens in the temporal window $[t-N, t]$.

In the following let us concentrate on a single node and a given k-steps neighborhood; for notational brevity we drop all super-scripts (i) and sub-scripts k.

Change detection

Consider the following set of hypotheses for detecting disconnection-like changes over a window of $N+1$ epochs

$$\begin{cases} \mathcal{H}_0 : & S(t-N) = \ldots = S(t-T) = \overline{S}, \\ & S(i) \geq \sigma \overline{S} \text{ for all } i \in \{t-T+1, \ldots, t\} \\ \mathcal{H}_1 : & S(t-N) = \ldots = S(t-T) = \overline{S}, \\ & \text{ exists } i \in \{t-T+1, \ldots, t\} \text{ s.t. } S(i) < \sigma \overline{S} \end{cases} \tag{1}$$

- $\mathcal{H}_0$ assumes that before the change time, $t-T$, the true size is constant and equal to $\overline{S}$; after the change the network size ‘remains greater than‘ $\sigma\overline{S}$. In particular the size could have remained constant.
- $\mathcal{H}_1$ assumes again that before the change time the true size is equal to $\overline{S}$; in this case, however, a disconnection happened after $t-T$.
- $\sigma \in (0,1]$ tunes the test sensibility to small size disconnections

Change detection

There are no nominal parameters: the pre-change size, $\overline{S}$, the change time, $t - T$, and the post-change sizes $S(\tau)$, $\tau = t - T + 1, \ldots, t$ are unknowns that must be estimated.

Since we assume no a-priori knowledge on them, these quantities are estimated from the sample $\boldsymbol{f}(t - N, t)$ employing ML approaches.

Remark: $\overline{S}$ and T define the set of hypotheses to be tested. Once they are estimated they are treated as deterministic parameters by the actual decision procedure. Apparently, this not fully rigorous.

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Alg. Neighborhood change detection

1) *(cycle on all the plausible change times)*

for $\mathcal{T} = 1, \ldots, N-1$ **do**

2) *(estimation of the pre-change value)*

$$\overline{S}(\mathcal{T}) = \frac{M(N-\mathcal{T}+1)}{\sum_{\tau=t-N}^{t-\mathcal{T}} \sum_{m=1}^{M} -\log\left(f_m(\tau)\right)}$$

---

3) *(estimation of the post-change values)*

for $\tau \in \{t-\mathcal{T}+1, \ldots, t\}$ **do**

$$\widehat{S}(\tau) = \frac{M}{\sum_{m=1}^{M} -\log\left(f_m(\tau)\right)}$$

$$\widehat{S}_0(\tau) = \begin{cases} \widehat{S}(\tau) & \text{if } \widehat{S}(\tau) \geq \sigma \overline{S}(\mathcal{T}) \\ \sigma \overline{S}(\mathcal{T}) & \text{otherwise} \end{cases}$$

end for

4) *(computation of the GLR)*

$$\Lambda(\mathcal{T}) = \frac{\prod_{\tau=t-\mathcal{T}}^{t} \ell\left(\widehat{S}_0(\tau)\,;\ \boldsymbol{f}(\tau)\right)}{\prod_{\tau=t-\mathcal{T}}^{t} \ell\left(\widehat{S}(\tau)\,;\ \boldsymbol{f}(\tau)\right)}$$

end for

---

5) *(computation of the optimal change time)*

$$T = \arg \min_{\mathcal{T}\in\{1,\dots,N-1\}} \Lambda(\mathcal{T})$$

---

6) *(decision)*

$$g\left(\boldsymbol{f}(t-N:t)\right) = \begin{cases} 0 & \text{if } \Lambda(T) \geq \lambda \\ 1 & \text{otherwise.} \end{cases}$$

Change detection

In the report we proved that once $\overline{S}$ and T are fixed the following facts hold:

- the power function establishes a partial order in parameter space: if $S(\tau) \leq S'(\tau)$ for all $\tau = t - T + 1, \ldots, t$, then

$$\beta_g \left(S(t - T + 1), \ldots, S(t)\right) \geq \beta_g \left(S'(t - T + 1), \ldots, S'(t)\right) \ .$$

- the test size can be computed by evaluating the the power function at the point $S(\tau) = \sigma\overline{S}\ ,\ \tau = t - T + 1, \ldots, t$, i.e.

$$\alpha_0 = \beta_g(S(\tau) = \sigma\overline{S}\ ,\ \tau = t - T + 1, \ldots, t)$$

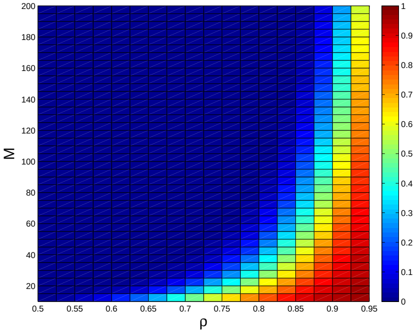
- the distribution of the GLR depends only on the ratios $\rho(\tau) = S(\tau)/(\sigma\overline{S}, \tau = t - T + 1, \ldots, t$. In particular, the distribution of the GLR used in the computation of α_0 does not depend on the outcome of $\overline{S}$.

Change detection

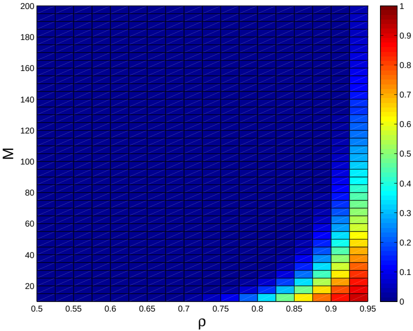
It follows that nodes do not need to compute the test threshold, λ, at each epoch after $\overline{S}$ and T have been estimated. N thresholds, one for each possible outcome of $T \in \{1, \ldots, N\}$, can be precomputed off-line, stored in the node's memory and then used at run-time.

The GLR distribution (for each T) can be expressed in analytical form by means of Lambert's W function. This part, however, is missing from the report due to lack of time. Instead we evaluated it using the Monte Carlo method.

Change detection



(a) $T = 5$



(b) $T = 20$

Figure: Type II error probability as a function of M, T and $\rho := S(\tau)/\sigma\overline{S}$, $\tau = t - T + 1, \ldots, t$; the test size is $\alpha_0 = 0.01$.

Simulations

We have ran extensive simulations of the change detection algorithm on a prototypical grid network with 100 nodes. We wanted to

- analyze the role of the tunable parameters
- understand if the change detectors could be used to ‘identify‘ the direction into which the disconnection happened, e.g. by distributedly compute a gradient towards the disconnected sub-graph

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Simulations - Static Network

Let us start with the following choice of parameters

$$M = 50,\ D = 20,\ N = 25,\ \alpha_0 = 0.01,\ \sigma = 1\ .$$

Simulations - Static Network

The false-positive rate is very high, greater that what one would expect from setting $\alpha_0 = 0.01$.

The pre-change size estimate, $\overline{S}$, tends to be very noise and biased towards over estimating the true value.

We believe this is due to the näive ML approach for estimating the change time, $t - T$, that we employ. This estimator tends to be 'tricked' by outliers in $\widehat{S}_k(t)$ at times near the lower-end of temporal window $[t - N, t]$.

One possible solution is to restrict the search for $t - T$ to the subset $[t - N + \overline{N}, t]$. Thus reserving $\overline{N}$ epochs for the sole computation of the pre-change size.

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Simulations - Static Network

One possible solution is to restrict the search for $t - T$ to a subset $[t - N + \overline{N}, t]$ of the temporal window. Thus reserving $\overline{N}$ epochs for the sole computation of the pre-change size.

Simulations - Dynamic Network

We model a dynamic network in which agents are subject to failure by allowing each node to be in one of two states: active or inactive.
At each epoch active nodes have a probability $P_d = 0.01$ of transitioning to the inactive state while nodes currently inactive have a probability $P_b = 0.04$ to transition back to the active state.

Simulations - Dynamic Network

The rate of false-positives has increased with respect to the previous case. This is again due to over estimation of $\overline{S}$.

This is the test case that motivated the introduction of the parameter σ in the definition of $\mathcal{H}_0$ and $\mathcal{H}_1$.

Simulations - Distance to disconnected sub-graph

Information from the change detectors can be exploited to estimate one node's distance from the disconnected sub-graph. Let us define for each node $i \in \mathcal{N}(t)$

$$\mathcal{K}^{(i)}(t) := \Big\{ k \; : \; i \text{ detected a change in its } k\text{-steps} \\ \text{neighborhood at time } t, \; k = 1, \ldots, D \Big\} .$$

Then one way of evaluating how far a node is (in hops/epochs) from the disconnection is by employing the estimator

$$\widehat{\delta}^{(i)}(t) := \min_{j \in \mathcal{N}(t)} \left\{ \mathcal{K}^{(j)}(t) \right\} .$$

$\delta^{(i)}(t)$ is the smallest k for which node i detected a change in its k-steps neighborhoods.

Simulations - Distance to disconnected sub-graph

In principle $\delta^{(i)}(t)$ could be used to distribuitedly build a gradient towards the disconnection

Simulations - Size preserving topology changes

Our approach, by design, can only detect changes in the size of a node's neighborhood but not on its ‘connectivity‘.