



HAL
open science

Maximum Coverage and Maximum Connected Covering in Social Networks with Partial Topology Information

Patricio Reyes, Alonso Silva

► **To cite this version:**

Patricio Reyes, Alonso Silva. Maximum Coverage and Maximum Connected Covering in Social Networks with Partial Topology Information. 6th IEEE International Workshop on Network Science for Communication Networks, May 2014, Toronto, Canada. hal-00922517

HAL Id: hal-00922517

<https://inria.hal.science/hal-00922517>

Submitted on 27 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Maximum Coverage and Maximum Connected Covering in Social Networks with Partial Topology Information

Patricio Reyes ^{*1} and Alonso Silva ^{†2}

¹ Dept. Statistics, Universidad Carlos III de Madrid (UC3M), Calle Madrid 126, UC3M - Estadística, 28903 Getafe, Madrid, Spain

² Alcatel-Lucent Bell Labs France, Centre de Villarceaux, Route de Villejust, 91620 Nozay, France

Abstract

Viral marketing campaigns seek to recruit the most influential individuals to cover the largest target audience. This can be modeled as the well-studied maximum coverage problem. There is a related problem when the recruited nodes are connected. It is called the maximum connected cover problem. This problem ensures a strong coordination between the influential nodes which are the backbone of the marketing campaign. In this work, we are interested on both of these problems. Most of the related literature assumes knowledge about the topology of the network. Even in that case, the problem is known to be NP-hard. In this work, we propose heuristics to the maximum connected cover problem and the maximum coverage problem with different knowledge levels about the topology of the network. We quantify the difference between these heuristics and the local and global greedy algorithms.

1 Introduction

One of the main objectives of viral marketing campaigns is to find the most influential individuals to cover the largest target audience. This problem can be modeled as the well-studied maximum coverage

problem. In the need of coordination through the marketing campaign, a more relevant objective is to seek the most influential connected individuals. Hereby the connectedness will be fundamental since the advertisement needs to spread quickly through the network. In this work, we are interested on both of these problems. Most of the related works on these topics assume knowledge about the topology of the network. Even in that case, the problem is known to be NP-hard. Recently, in [1] the authors propose a (local) greedy algorithm to the maximum connected covering problem by learning the topology of the network on-the-fly.

In this work, we present different heuristics to both of these problems with different levels of knowledge about the topology of the network. We quantify the difference between these algorithms. Obviously, different knowledge about the topology of the network will restrict us to use different heuristics with the problem at hand.

Works providing heuristics to maximize the impact of a virus marketing campaign are [2] and [3]. Other works have been interested on the spreading of information through cascades on a weighted influence graph and some internal conviction threshold of the individuals (see e.g. [4, 5, 6]). Their work is based on the submodularity of the local spreading and the bounds on the performance of the greedy algorithms for submodular functions given in [7]. Our work is

*Email: preyes@est-econ.uc3m.es

†Email: alonso.silva@alcatel-lucent.com

different since we are not interested on the spreading of information through cascades, but on a “quick” spreading: we are interested on a one-hop spread of information only through neighbors. Closely related works are [1] and [8], where the authors assume a one-hop lookahead [1] and two-hops lookaheads [8].

The paper is structured as follows. In Section 2, we formulate both, the maximum coverage problem and the maximum connected covering problem. In Section 3, we describe the different levels of knowledge about the topology of the network that we consider in this work. In Section 4, we present existing heuristics and we present some new heuristics to these problems based on the different levels of knowledge of the network. In Section 5, we present the simulations of the algorithms and finally we conclude in Section 6.

2 Problem Formulation

We consider an influence graph $G = (V, E)$ where V is the set of vertices and $E \subseteq V \times V$ is the set of edges. Each vertex of the graph represents an individual and each edge represents a relationship of mutual influence between them (e.g. friendship over a social network). An individual $i \in V$ has influence over another individual $j \in V$ if and only if $\{i, j\} \in E$. We assume that the influence graph G is an undirected graph with no self-loops. We denote by $\mathcal{N}(i)$ the set of neighbors of vertex i , i.e., $\mathcal{N}(i) = \{j \in V : \{i, j\} \in E\}$, and for a set of vertices $A \subseteq V$, we denote by $\mathcal{N}(A)$ the set of neighbors of A as $\mathcal{N}(A) = \{j \in V \setminus A : \text{exists } i \in A \text{ such that } \{i, j\} \in E\}$.

We consider that time is slotted, i.e., $t \in \mathbb{N} \cup \{0\}$. We denote by $\mathcal{R}(t)$ the set of recruited individuals at time $t \geq 0$. In particular, $\mathcal{N}(\mathcal{R}(t))$ is the set containing unrecruited neighbors of $\mathcal{R}(t)$.

The algorithms that we present are sequential algorithms which proceed as follows: at time t , with $0 \leq t \leq K$, the algorithm recruits a node $i \in V \setminus \mathcal{R}(t-1)$ and performs the update $\mathcal{R}(t) = \mathcal{R}(t-1) \cup \{i\}$.

The objective of the maximum coverage algorithms is to maximize the size of the network covering $\mathcal{C}(t) = \mathcal{R}(t) \cup \mathcal{N}(\mathcal{R}(t))$ and in the case of the maximum connected covering problem this objective is subject to

the additional constraint that the set $\mathcal{R}(t)$ must be connected.

The degree $d(i)$ of a node $i \in V$ is the number of neighbors of a node, i.e., $d(i) = |\mathcal{N}(i)|$ where $|\cdot|$ is the cardinality function. The observed degree $d_{obs}(i, t)$ of a node $i \in V$ at time t is the number of recruited neighbors or neighbors of recruited neighbors of i , i.e., $d_{obs}(i, t) = |\{j \in \mathcal{R}(t) \cup \mathcal{N}(\mathcal{R}(t)) : \{i, j\} \in E\}|$. The excess degree $d_{excess}(i, t)$ of a node $i \in V$ at time t is difference between the degree and the observed degree of node i at time t , i.e., $d_{excess}(i, t) = d(i) - d_{obs}(i, t)$.

3 Information Levels

For both of the problems we are dealing with in this work, we consider different levels of information about the topology of the network.

1. List of nodes: we consider that the recruiter knows the list of nodes (we know V) so there is a knowledge about the nodes the network has and there is a possibility to recruit any node within the network. Once a node has been recruited we consider that the recruited node gives information about who are its neighbors.
2. One-hop lookahead: we consider that the recruiter knows only one node, denoted i , and once a node is recruited it gives information about who are its neighbors and who are their mutual neighbors (between recruited nodes). Actually, the recruiter may only need to know the quantity of neighbors, observed neighbors and mutual neighbors (between recruited nodes), in order to compute the excess degree.
3. Two-hops lookahead: we consider that the recruiter knows only one node, denoted i , and once a node is recruited it gives information about who are its neighbors and neighbors of neighbors and who are their neighbors and the mutual neighbors.
4. List of nodes and two-hops lookahead: we have knowledge about the list of nodes as in 1) and two-hops lookahead as in 3).

5. Full knowledge: we consider that the recruiter has full knowledge about the topology of the network. It knows the set of nodes V and the set of edges E .

We notice that in [1] the knowledge level that they consider is 2) since in their case, they do not have any information about the network topology and they are discovering the network while they are recruiting over the network.

4 Description of Algorithms

In this section, we give a brief description of the algorithms for the different scenarios (levels of information) in both problems: the maximum coverage problem (SCP) and maximum connected covering (MCC) problem.

4.1 Set Covering Problem (SCP)

In the first scenario, called SCP 1, we consider that the recruiter knows the list of nodes but doesn't have any information about the topology of the graph as in 3 1). Once we recruit a node and only then, we consider that the node gives us information about which nodes it is connected to. Under these characteristics, we consider Algorithm 1. Given that initially you don't have any information about the topology of the network, Algorithm 1 simply chooses to recruit a node at random and since then the recruiter knows to which nodes it is connected to, it can remove those nodes (since they are already covered) from the uncovered nodes and then again choose a node from within the set of remaining uncovered nodes at random.

For the probability distribution over a set of nodes $S \subseteq V$, we identify each node $i \in S$ with a unique integer from 1 to $|S|$. We consider a probability distribution ζ over the set of nodes $|S|$, i.e., $\zeta(i) \geq 0$ and $\sum_{i \in S} \zeta(i) = 1$. For simplicity, we consider the two following cases:

- The uniform distribution $\zeta_1(i) = 1/|S|$,
- The degree distribution $\zeta_2(i) = d(i)/\sum_{j \in V} d(j)$.

However, we notice that the probability distribution ζ is not restricted to these two choices.

Algorithm 1 SCP 1: Random

- 1: Initialize the list of uncovered nodes U with the set of all nodes $U \leftarrow V$, the list of recruited nodes R with the empty set $R \leftarrow \emptyset$, and the list of covered nodes C with the empty set $C \leftarrow \emptyset$,
 - 2: $k \leftarrow 1$,
 - 3: **repeat**
 - 4: Recruit a node $i \in U$ uniformly at random, i.e., $R \leftarrow R \cup \{i\}$,
 - 5: Remove node i and its neighbors $\mathcal{N}(i)$ from the list of uncovered nodes, i.e., $U \leftarrow U \setminus (i \cup \mathcal{N}(i))$
 - 6: Add node i and its neighbors $\mathcal{N}(i)$ to the list of covered nodes, i.e., $C \leftarrow C \cup (i \cup \mathcal{N}(i))$
 - 7: $k \leftarrow k + 1$,
 - 8: **until** $k > K$ or $U \leftarrow \emptyset$
-

In the second scenario, called SCP 2, we assume that when a node is recruited it provides a two-hops lookahead information, i.e., it gives information about its neighbors and the neighbors of its neighbors as in 3 3). To take advantage of this knowledge, Algorithm 2 which was originally proposed by Guha and Khuller [8], proposes to recruit the node from within a two-hop neighborhood that have the maximum number of uncovered neighbors (maximize the excess degree), and then again to choose the node from within a two-hop neighborhood of the set of recruited nodes that have the maximum number of uncovered neighbors.

In the third scenario, called SCP 3, we consider that the recruiter knows the list of nodes (as in the first scenario) and that when a node is recruited it provides a two-hop lookahead information (as in the second scenario). To take advantage of this knowledge, we propose Algorithm 3 that at every step with probability δ recruits a node at random from within the set of uncovered nodes and with probability $(1-\delta)$ recruits the node from within a two-hop neighborhood that have the maximum number of uncovered neighbors. It is clear that the appeal from this version of the algorithm is that it is a probabilistic combination from both previous scenarios.

Algorithm 2 SCP 2: Two-hops Greedy Algorithm[8]

- 1: Initialize the list of uncovered nodes U with the set of all nodes $U \leftarrow V$, the list of recruited nodes R with the empty set $R \leftarrow \emptyset$, and the list of covered nodes C with the empty set $C \leftarrow \emptyset$,
 - 2: $k \leftarrow 1$,
 - 3: **repeat**
 - 4: Recruit a node $i \in U \cap [\mathcal{N}(R) \cup \mathcal{N}(\mathcal{N}(R))]$ of maximum excess degree, i.e., $R \leftarrow R \cup \{i\}$ where i is such that $|\mathcal{N}(i) \setminus (R \cup \mathcal{N}(R))|$ is maximum restricted to the set $U \cap [\mathcal{N}(R) \cup \mathcal{N}(\mathcal{N}(R))]$,
 - 5: Remove node i and its neighbors $\mathcal{N}(i)$ from the list of uncovered nodes, i.e., $U \leftarrow U \setminus (i \cup \mathcal{N}(i))$,
 - 6: Add node i and its neighbors $\mathcal{N}(i)$ to the list of covered nodes, i.e., $C \leftarrow C \cup (i \cup \mathcal{N}(i))$
 - 7: $k \leftarrow k + 1$
 - 8: **until** $k > K$ or $U \leftarrow \emptyset$
-

For the probability distribution over a set of nodes $S \subseteq V$, we consider ζ as in the first scenario. We consider α to be a variable to be chosen $0 \leq \alpha \leq 1$.

The fourth scenario, called SCP 4, is the full knowledge scenario as in 3 5 where you know the topology of the network (the list of nodes, the list of neighbors of the nodes, the list of neighbors of the neighbors of the nodes, etc). Algorithm 4 chooses at each step greedily the node that have the maximum number of uncovered neighbors from the full set of uncovered nodes.

4.2 Maximum Connected Coverage (MCC) problem

In the first scenario, called MCC 1, we consider that we know a node, denoted node $i \in V$, and we consider that when a node is recruited it gives a one-hop lookahead as in 3 2). In Algorithm 5, we propose a random selection over the set of neighbors of the recruited nodes which are not themselves already recruited, i.e., $P = \mathcal{N}(R) \setminus R$. We notice that this scenario is different from a random walk since we are

Algorithm 3 SCP 3: THG + Random α

- 1: Initialize the list of uncovered nodes U with the set of all nodes $U \leftarrow V$, the list of recruited nodes R with the empty set $R \leftarrow \emptyset$, and the list of covered nodes C with the empty set $C \leftarrow \emptyset$,
 - 2: $k \leftarrow 1$,
 - 3: **repeat**
 - 4: Draw a Bernoulli random variable X with parameter α
 - 5: **if** $X = 1$ **then**
 - 6: Recruit a node $j \in U$ at random (according to ζ) from the set U , i.e., $R \leftarrow R \cup \{j\}$
 - 7: Remove node j and its neighbors $\mathcal{N}(j)$ from the list of uncovered nodes, i.e., $U \leftarrow U \setminus (j \cup \mathcal{N}(j))$
 - 8: Add node j and its neighbors $\mathcal{N}(j)$ to the list of covered nodes, i.e., $C \leftarrow C \cup (j \cup \mathcal{N}(j))$
 - 9: **else**
 - 10: Recruit a node $i \in U \cap [\mathcal{N}(R) \cup \mathcal{N}(\mathcal{N}(R))]$ of maximum excess degree, i.e., $R \leftarrow R \cup \{i\}$ where i is such that $|\mathcal{N}(i) \setminus (R \cup \mathcal{N}(R))|$ is maximum restricted to the set $U \cap [\mathcal{N}(R) \cup \mathcal{N}(\mathcal{N}(R))]$,
 - 11: Remove node i and its neighbors $\mathcal{N}(i)$ from the list of uncovered nodes, i.e., $U \leftarrow U \setminus (i \cup \mathcal{N}(i))$
 - 12: Add node i and its neighbors $\mathcal{N}(i)$ to the list of covered nodes, i.e., $C \leftarrow C \cup (i \cup \mathcal{N}(i))$
 - 13: **end if**
 - 14: $k \leftarrow k + 1$
 - 15: **until** $k > K$ or $U \leftarrow \emptyset$
-

Algorithm 4 SCP 4: Greedy Algorithm [7]

- 1: Initialize the list of uncovered nodes U with the set of all nodes $U \leftarrow V$, the list of recruited nodes R with the empty set $R \leftarrow \emptyset$, and the list of covered nodes C with the empty set $C \leftarrow \emptyset$,
 - 2: $k \leftarrow 1$,
 - 3: **repeat**
 - 4: Recruit a node $i \in U$ that maximizes the excess degree, i.e., $R \leftarrow R \cup \{i\}$, where $i \in U$ is such that $|\mathcal{N}(i) \setminus (R \cup \mathcal{N}(R))|$ is maximum,
 - 5: Remove node i and its neighbors $\mathcal{N}(i)$ from the list of uncovered nodes, i.e., $U \leftarrow U \setminus (i \cup \mathcal{N}(i))$,
 - 6: Add node i and its neighbors $\mathcal{N}(i)$ to the list of covered nodes, i.e., $C \leftarrow C \cup (i \cup \mathcal{N}(i))$
 - 7: $k \leftarrow k + 1$,
 - 8: **until** $k > K$ or $U \leftarrow \emptyset$
-

choosing among the whole set P and not only the neighbors of the newly recruited node.

In the second scenario, called MCC 2, we also consider that we know a node, denoted node $i \in V$, and we consider that when a node is recruited it gives the list of neighbors of the recruited nodes. In Algorithm 6, which was originally proposed by [1], the algorithm greedily recruits the node in P which maximizes the excess degree.

5 Simulations

We performed simulations of the previously described algorithms in Erdos-Renyi graphs $G(N, p_N)$ where $N \in \{50, 100, 150, 200, 250\}$ is the number of nodes in the graph and p_N is the probability of two nodes being connected. We chose $p_N = 2 \ln(N)/N$ to ensure connectivity. We simulated 30 instances for each graph size. In order to avoid problems for randomly choosing the initial node, we set 3 initial nodes (10 instances each) for each algorithm. In summary, each algorithm was run 3 times in each graph in each instance, starting by 3 different nodes. Therefore, the figures show the average of the number of recruited nodes (over all the graph instances with same size) needed to cover the whole graph. We notice that this

Algorithm 5 MCC 1: Random Neighbor

- 1: Initialize the list of uncovered nodes U with the set of all nodes $U \leftarrow V$, the list of recruited nodes R with the empty set $R \leftarrow \emptyset$, and the list of covered nodes C with the empty set $C \leftarrow \emptyset$,
 - 2: Recruit a node $i \in U$ at random (according to ζ), i.e., $R \leftarrow R \cup \{i\}$,
 - 3: Remove node i and its neighbors $\mathcal{N}(i)$ from the list of uncovered nodes, i.e., $U \leftarrow U \setminus (i \cup \mathcal{N}(i))$,
 - 4: Add node i and its neighbors $\mathcal{N}(i)$ to the list of covered nodes, i.e., $C \leftarrow C \cup (i \cup \mathcal{N}(i))$,
 - 5: Initialize the list of candidates to be recruited with the set of neighbors of i , i.e., $P \leftarrow \mathcal{N}(i)$,
 - 6: $k \leftarrow 2$
 - 7: **repeat**
 - 8: Recruit a node $j \in P$ uniformly at random from the set P , i.e., $R \leftarrow R \cup \{j\}$ with $j \in P$,
 - 9: Remove node j from the list of candidates to be recruited, i.e., $P \leftarrow P \setminus \{j\}$,
 - 10: Remove the node j and its neighbors $\mathcal{N}(j)$ from the list of uncovered nodes, i.e., $U \leftarrow U \setminus (j \cup \mathcal{N}(j))$,
 - 11: Add node j and its neighbors $\mathcal{N}(j)$ to the list of covered nodes, i.e., $C \leftarrow C \cup (j \cup \mathcal{N}(j))$,
 - 12: Add the unrecruited neighbors of j to the list of candidates to be recruited, i.e., $P \leftarrow P \cup (\mathcal{N}(j) \cap U)$,
 - 13: $k \leftarrow k + 1$
 - 14: **until** $k > K$ or $U \leftarrow \emptyset$
-

Algorithm 6 MCC 2: Online Myopic MCC [1]

- 1: Initialize the list of uncovered nodes U with the set of all nodes $U \leftarrow V$, the list of recruited nodes R with the empty set $R \leftarrow \emptyset$, and the list of covered nodes C with the empty set $C \leftarrow \emptyset$,
 - 2: Recruit node $i \in U$, i.e., $R \leftarrow R \cup \{i\}$,
 - 3: Remove node i and its neighbors $\mathcal{N}(i)$ from the list of uncovered nodes, i.e., $U \leftarrow U \setminus (i \cup \mathcal{N}(i))$,
 - 4: Add node i and its neighbors $\mathcal{N}(i)$ to the list of covered nodes, i.e., $C \leftarrow C \cup (i \cup \mathcal{N}(i))$,
 - 5: $k \leftarrow 2$
 - 6: **repeat**
 - 7: Recruit a node $i \in U$ that maximizes the excess degree, i.e., $R \leftarrow R \cup \{i\}$, where $i \in U$ is such that $|\mathcal{N}(i) \setminus (R \cup \mathcal{N}(R))|$ is maximum,
 - 8: Activate a node $i \in U$ that maximizes the excess degree, i.e., $R = R \cup \{i\}$, where $i \in U$ is such that $|\mathcal{N}(R) \cap \mathcal{N}(i)|$ is maximum,
 - 9: Activate one of the nodes $i \in U$ of maximum excess degree, i.e., $R = R \cup \{i\}$ where i is such that $d_i - d_i^{obs} = \max_{k \in \{1, \dots, n\}} d_k - d_k^{obs}$ where d^{obs} is the observed degree.
 - 10: Remove the node i and its neighbors $\mathcal{N}(i)$ from the list of uncovered nodes, i.e., $U = U \setminus (i \cup \mathcal{N}(i))$
 - 11: $k \leftarrow k + 1$
 - 12: **until** $k > K$ or $U = \{\emptyset\}$
-

corresponds to the case when there is no restriction over K but only on the number of uncovered nodes.

It is difficult to compare different knowledge levels since for example how can we compare between having the possibility of recruiting any node in a network but completely at random compared to be able to connect only to two hops away nodes but knowing exactly how many observed neighbors and neighbors do they have. The first observation that we can make of Figure 1 is that recruiting nodes at random SCP 1 performs 56% worst than having a two-hops lookahead and a greedy algorithm SCP 3 ($((\text{SCP 1} - \text{SCP 3})/\text{SCP 3}) \times 100$). The second observation which we found surprising was that in Erdos-Renyi graphs the greedy approach works better than the mixed approach (algorithm SCP 3 which combines the greedy approach and the random choice). The reason why we were expecting to have a different behavior is because the algorithm may start in a bad initial location and through a greedy approach it may take a while before finding good nodes to be recruited. In fact, SCP 3 performs 17% worst than the greedy approach SCP 3 ($((\text{SCP 2} - \text{SCP 3})/\text{SCP 3}) \times 100$). We believe that the performance of SCP 3 may improve by modifying the parameter α which we took as $\alpha = 1/2$.

We performed simulations of the previously described algorithms also in Barabasi-Albert graphs. The chosen Barabasi-Albert graphs were undirected graphs and were generated as follows. We started with a single vertex. At each time step, we added one vertex and the new vertex connects two edges to the old vertices. The probability that an old vertex is chosen is proportional to its degree.

In Figure 3 (similarly to Figure 1), we notice that recruiting nodes at random SCP 1 performs 196% worst than having a two-hops lookahead and a greedy algorithm SCP 3 ($((\text{SCP 1} - \text{SCP 3})/\text{SCP 3}) \times 100$). The mixed approach with combines the greedy approach and the random choice SCP 2 performs 19% worst than the greedy approach SCP 3 ($((\text{SCP 2} - \text{SCP 3})/\text{SCP 3}) \times 100$).

Similarly, in Figure 4, we have that to choose uniformly at random between uncovered nodes performs very poorly compared to the greedy one-hop lookahead algorithm.

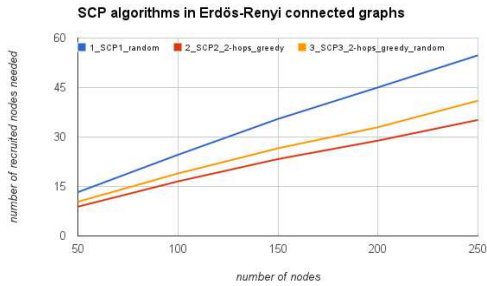


Figure 1: Number of recruited nodes needed vs number of nodes

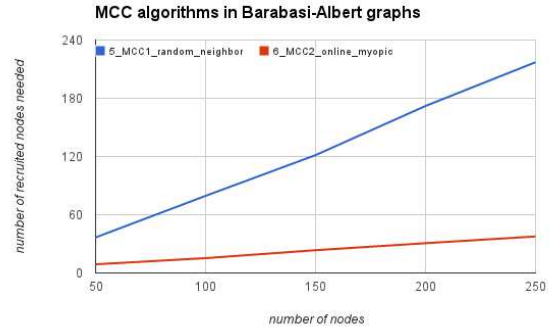


Figure 4: Number of recruited nodes needed vs number of nodes

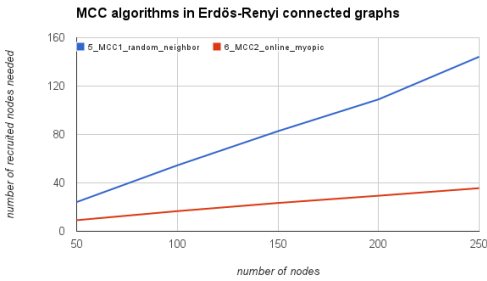


Figure 2: Number of recruited nodes needed vs number of nodes

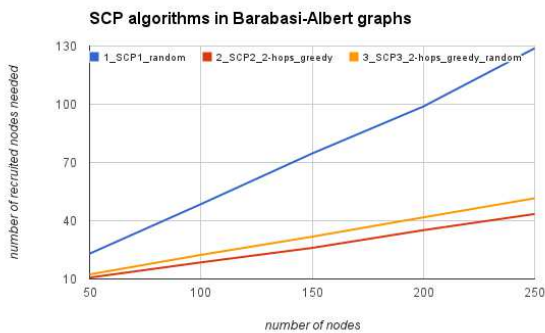


Figure 3: Number of recruited nodes needed vs number of nodes

6 Conclusions and Future Directions

In this work, we were interested on two different problems: the maximum coverage problem and the maximum connected covering problem. The motivation of our work is viral marketing campaigns on social networks. Our perspective was to analyze both problems from the knowledge we may have of the topology of the network. We presented some existing and new heuristics to both of these problems. We quantified how different levels of information have an effect on the type of algorithm that we choose and this translates on a better or worst performance depending on the knowledge we have on the topology of the network.

There are many interesting future directions to this work. Just to name a few, one direction is to provide theoretical bounds to the new heuristics and to consider digraphs instead of undirected graphs. Another direction is to study how changes on the topology of the network can affect the problem at hand. Moreover, if there are changes constantly, how to make the maximum coverage set and maximum connected covering set to change together with this dynamicity.

Acknowledgments

The work of A. Silva was partially carried out at LINCS (www.lincs.fr).

References

- [1] K. Avrachenkov, P. Basu, G. Neglia, B. F. Ribeiro, and D. Towsley, “Online myopic network covering,” *CoRR*, vol. abs/1212.5035, 2012.
- [2] P. Domingos and M. Richardson, “Mining the network value of customers,” in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD ’01, (New York, NY, USA), pp. 57–66, ACM, 2001.
- [3] M. Richardson and P. Domingos, “Mining knowledge-sharing sites for viral marketing,” in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD ’02, (New York, NY, USA), pp. 61–70, ACM, 2002.
- [4] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-03)*, (New York), pp. 137–146, ACM Press, Aug. 24–27 2003.
- [5] D. Kempe, J. Kleinberg, and É. Tardos, “Influential nodes in a diffusion model for social networks,” in *ICALP: Annual International Colloquium on Automata, Languages and Programming*, 2005.
- [6] Mossel and Roch, “On the submodularity of influence in social networks,” in *STOC: ACM Symposium on Theory of Computing (STOC)*, 2007.
- [7] G. Nemhauser, L. Wolsey, and M. Fisher, “An analysis of approximations for maximizing submodular set functions I,” *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.
- [8] S. Guha and S. Khuller, “Approximation algorithms for connected dominating sets,” *Algorithmica*, vol. 20, 1996.