



Effects of Network Communities and Topology Changes in Message-Passing Computation of Harmonic Influence in Social Networks

Wilbert Samuel Rossi, Paolo Frasca

► To cite this version:

Wilbert Samuel Rossi, Paolo Frasca. Effects of Network Communities and Topology Changes in Message-Passing Computation of Harmonic Influence in Social Networks. CPHS 2018 - 2nd IFAC Conference on Cyber-Physical & Human Systems, Dec 2018, Miami, United States. pp.1-4. hal-01877617

HAL Id: hal-01877617

<https://hal.archives-ouvertes.fr/hal-01877617>

Submitted on 20 Sep 2018

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.

Effects of Network Communities and Topology Changes in Message-Passing Computation of Harmonic Influence in Social Networks

W. S. Rossi * P. Frasca **

* *University of Twente, 7500 AE Enschede, The Netherlands (e-mail: w.s.rossi@utwente.nl).*

** *Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-lab, 38000 Grenoble, France (e-mail: paolo.frasca@gipsa-lab.fr).*

Abstract: The harmonic influence is a measure of the importance of nodes in social networks, which can be approximately computed by a distributed message-passing algorithm. In this extended abstract we look at two open questions about this algorithm. How does it perform on real social networks, which have complex topologies structured in communities? How does it perform when the network topology changes while the algorithm is running? We answer these two questions by numerical experiments on a Facebook ego network and on synthetic networks, respectively. We find out that communities can introduce artefacts in the final approximation and cause the algorithm to overestimate the importance of “local leaders” within communities. We also observe that the algorithm is able to adapt smoothly to changes in the topology.

Keywords: Distributed algorithm, Message-passing, Opinion dynamics, Social networks

1. HARMONIC INFLUENCE AND MESSAGE PASSING

In the study of social networks and dynamical processes therein, it is important to identify the most influential leaders. Several definitions have been used to evaluate nodes as potential leaders, e.g. Lin et al. (2014); Fitch and Leonard (2016); Van Mieghem et al. (2017). The *harmonic influence* is a definition motivated by a linear opinion dynamics model with stubborn agents. It was introduced in Vassio et al. (2014) and implicitly used in Acemoglu et al. (2013); Yildiz et al. (2013). We recall its equivalent definition given by Rossi and Frasca (2018). Consider a simple weighted graph¹ $\mathcal{G} = (I, E, C)$ with node set $I = \{\mathfrak{f}, 1, 2, \dots, n\}$ where \mathfrak{f} is a special node called *field*. The edge set E contains unordered pairs of nodes and the non-negative weight matrix $C \in \mathbb{R}_+^{I \times I}$ is such that C_{ij} and C_{ji} are both non-zero if and only if $\{i, j\} \in E$. We also introduce the diagonal matrix $D = \text{Diag}(C\mathbf{1})$ and the *Laplacian* matrix $L = D - C$. We assume matrix C to be symmetric and graph \mathcal{G} to be connected. Given a node $\ell \neq \mathfrak{f}$ where ℓ stands for *leader*, let $R^\ell := I \setminus \{\mathfrak{f}, \ell\}$ be the set of remaining nodes and consider the discrete Dirichlet problem

$$\begin{cases} (L\mathbf{x})_{R^\ell} = \mathbf{0} \\ x_\ell = 1 \\ x_{\mathfrak{f}} = 0. \end{cases} \quad (1)$$

¹ Vectors are denoted with boldface letters and matrices with capital letters. The all-zero and all-one vectors are denoted by $\mathbf{0}$ and $\mathbf{1}$, respectively. A graph is said to be connected if for any pair of nodes i, j there exists a sequence of adjacent edges that joins them.

The *harmonic influence* of ℓ is the sum of entries of the vector \mathbf{x} solution of (1), that is,

$$H(\ell) := \mathbf{1}^\top \mathbf{x}. \quad (2)$$

Then, the computation of the harmonic influence of the n nodes in $I \setminus \{\mathfrak{f}\}$ requires the solution of n linear systems. A naive approach would then require global knowledge of the graph and would not exploit apparent redundancies in the computations. To overcome these issues, Vassio et al. (2014) proposed the following distributed *Message Passing Algorithm* (MPA) that computes the influences of all nodes at the same time.

Let $t \in \{0, 1, \dots\}$ be an iteration counter and let the set $N_i = \{j \in I : \{i, j\} \in E\}$ contain the neighbors of i in \mathcal{G} . At each step, every node i sends to all its neighbors j two messages:

$$W^{i \rightarrow j}(t) \in [0, 1], \quad H^{i \rightarrow j}(t) \in [0, +\infty).$$

The field node \mathfrak{f} sends null messages:

$$W^{\mathfrak{f} \rightarrow j}(t) = 0, \quad H^{\mathfrak{f} \rightarrow j}(t) = 0, \quad \forall j \in N_{\mathfrak{f}}, \quad \forall t \geq 0,$$

whereas any other node $i \neq \mathfrak{f}$ sends the initial messages:

$$W^{i \rightarrow j}(0) = 1, \quad H^{i \rightarrow j}(0) = 1, \quad \forall j \in N_i \quad (3)$$

and then synchronously updates the messages sent to his neighbor j following the rules:

$$W^{i \rightarrow j}(t+1) = \left(1 + \sum_{k \in N_i^j} \frac{C_{ik}}{C_{ij}} (1 - W^{k \rightarrow i}(t))\right)^{-1} \quad (4)$$

$$H^{i \rightarrow j}(t+1) = 1 + \sum_{k \in N_i^j} W^{k \rightarrow i}(t) H^{k \rightarrow i}(t), \quad (5)$$

where $N_i^j := N_i \setminus \{j\}$ is the set of neighbors of i except the one to which the message is sent. At any time, any node ℓ in $I \setminus \{\mathfrak{f}\}$ can compute an approximation of $H(\ell)$ by

$$H^\ell(t) = 1 + \sum_{i \in N_\ell} W^{i \rightarrow \ell}(t) H^{i \rightarrow \ell}(t).$$

The MPA is exact on trees, where it converges in a number of steps equal to the diameter of the graph. On general graphs, the algorithm converges asymptotically as proved in Rossi and Frasca (2016, 2018). Based on extended simulations on random graphs, the typical convergence time of the algorithm is conjectured to be $O(m/n)$, where m is the number of edges. A mean-field argument by Rossi and Frasca (2017) corroborates this conjecture for homogeneous networks. In general, the limit values overestimate the exact values of the harmonic influences (that is, $H^\ell(\infty) \geq H(\ell)$), but the algorithm provides a ranking between the nodes that, on random graphs, is in very good agreement with the exact ranking.

In Section 2, we look at the correctness and convergence time for real social networks, which have a pronounced community structure. While convergence is guaranteed by the theoretical results, it not obvious whether that the convergence time follows the conjecture and whether the ranking remains meaningful. In Section 3, we extend the algorithm to networks that change topology while the MPA computation unfolds. For this case, convergence is not guaranteed by the available theory.

2. THE EFFECTS OF COMMUNITY STRUCTURE

Real social networks have complex topologies of interconnections that are often organized in communities. On one side, the degree distribution of real networks (i.e. the distribution of the number of interconnection of each node) is typically broad, with relatively few nodes of high degree and many of low degree. On the other side, nodes can be grouped in communities, such that most of the edges are concentrated within each community Fortunato (2010). In this section, we test the MPA on a real social network with community structure, extracted from the dataset² collected by Leskovec and Mcauley (2012).

The dataset contains a collection of *ego networks* from the Facebook social graph. Let $F = (U, A)$ be the full unweighed Facebook social graph at the data collection time, where U is the user set and A the edge set representing the acquaintance relations. Given a user $u \in U$, the set N_u is the set of Facebook friends of u . The ego network of u is the subgraph of F induced by N_u , i.e

$$(N_u, \{\{v, w\} \in A : v, w \in N_u\}),$$

and does not contain u . From the dataset we extracted one ego network with 885 nodes and 23960 edges. We identified the communities of the ego network with the ‘‘Louvain algorithm’’ by Blondel et al. (2008); in particular, we used the implementation³ by Rubinov and Sporns (2010) with default parameters. We found three prominent communities with 326, 434 and 125 nodes, see Fig. 1.

From the ego network we build two weighted graphs for the simulations. Graph $\mathcal{G}_1 = (I_1, E_1, C_1)$ contains all the ego network augmented by the field node. The node set is $I_1 = \{\mathfrak{f}, 1, \dots, 885\}$ and the edge set E_1 contains all the 23960 edges of the ego network. The edge set also include every edge of the form $\{\mathfrak{f}, i\}$ with $i \in \{1, \dots, 885\}$: we can interpret the field node \mathfrak{f} as the original user u and these

² <http://snap.stanford.edu/data/egonets-Facebook.html>

³ <https://sites.google.com/site/bctnet>

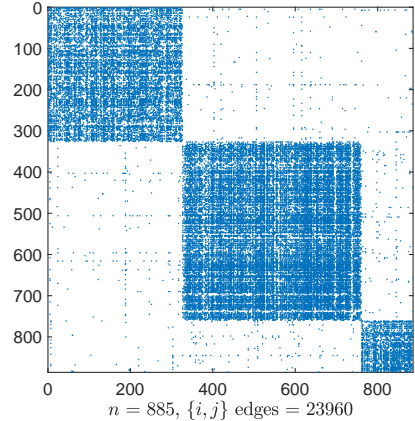


Fig. 1. The adjacency matrix of the selected ego network. The rows and columns are reordered according to the community structure, showing three prominent communities with 326, 434 and 125 nodes, respectively. A closer inspection reveals that within each community nodes with high and low degree coexist.

edges as those between u and his friends in N_u . Finally, the entries of the matrix $C_1 \in \mathbb{R}_+^{I_1 \times I_1}$ are

$$\begin{cases} (C_1)_{i\mathfrak{f}} = (C_1)_{\mathfrak{f}i} = 0.040 & \text{for every } i \in \{1, \dots, 885\} \\ (C_1)_{ij} = 1 & \text{if } i, j \neq \mathfrak{f} \text{ and } \{i, j\} \in E_1 \\ (C_1)_{ij} = 0 & \text{if } \{i, j\} \notin E_1 \end{cases}$$

Graph $\mathcal{G}_2 = (I_2, E_2, C_2)$ is the subgraph of \mathcal{G}_1 induced by $I_2 = \{\mathfrak{f}, 327, \dots, 760\}$. It is restricted to the second community of the ego network, augmented with the field node and his edges. The edge set E_2 contains 16253 edges between the 434 non-field nodes; the matrix C_2 follows accordingly.

We first discuss the simulation on graph \mathcal{G}_2 , with a single community but with nodes of very different degree. Fig. 2 represents the convergence of the MPA: the $W^{i \rightarrow j}(t)$ messages take about 20 steps to converge while the estimates $H^\ell(t)$ of the harmonic influence are very slow, taking almost 49000 iterations. Fig. 3 compares the estimates $H^\ell(\infty)$ with the exact values $H(\ell)$ computed with the definition. The MPA algorithm largely overestimates the harmonic influence, but the ranking remains well preserved.

The simulation on graph \mathcal{G}_1 presents interesting differences. Fig. 4 represents the convergence of the algorithm. The $W^{i \rightarrow j}$ messages take about 15 steps to converge while $H^\ell(t)$ take about 34000 steps. Since graph \mathcal{G}_1 is larger than \mathcal{G}_2 , but less dense, this moderate decrease in the convergence time is consistent with the $O(m/n)$ conjecture. Fig. 5 compares the estimates of the harmonic influence \mathcal{G}_1 with their exact counterpart; the crosses are colored according to the community of the nodes. The black crosses correspond to nodes in the (largest) second community, the blue crosses to nodes in the first community and the red ones to nodes in the third (and smallest) community. The community structure produces an interesting artefact, which is made apparent by the alignment of the crosses: the MPA assigns excess influence to leaders within smaller communities, compared to leaders of larger communities.

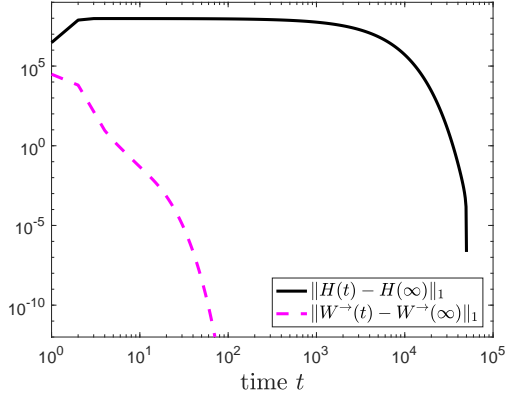


Fig. 2. The convergence of the MPA on the network \mathcal{G}_2 . The solid black line is the distance to convergence of the estimates of the harmonic influence obtained by the MPA. The dashed magenta line is the distance to convergence of the messages $W^{i \rightarrow j}(t)$.

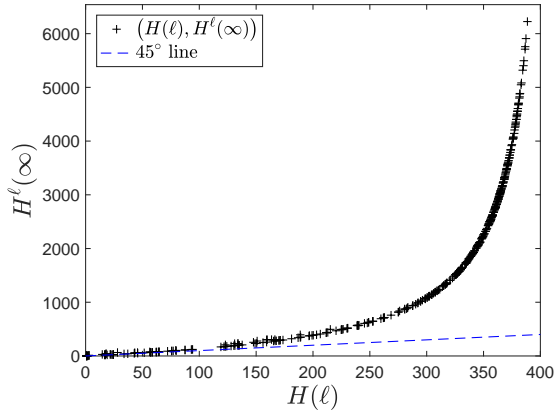


Fig. 3. The asymptotic values $H^\ell(\infty)$ of the harmonic influence computed by the MPA against the corresponding exact values $H(\ell)$ computed by the definition for graph \mathcal{G}_2 . All crosses are above the 45° line.

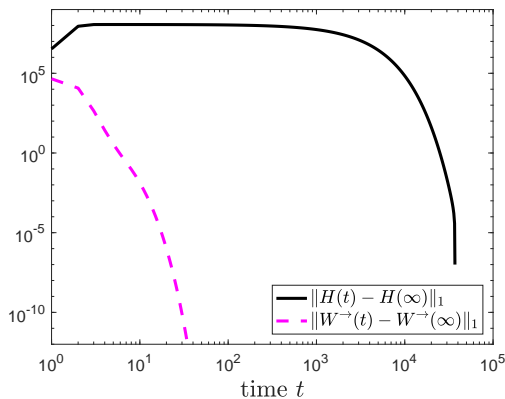


Fig. 4. The convergence of the MPA on the network \mathcal{G}_1 . The solid black line is the distance to convergence of the estimates of the harmonic influence obtained by the MPA. The dashed magenta line is the distance to convergence of the messages $W^{i \rightarrow j}(t)$.

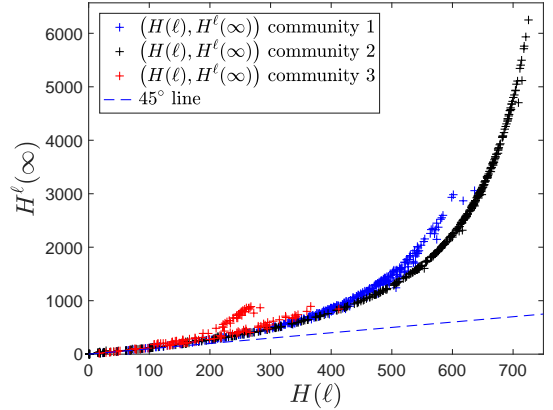


Fig. 5. The asymptotic values $H^\ell(\infty)$ of the harmonic influence computed by the MPA against the corresponding exact values $H(\ell)$ computed by the definition for graph \mathcal{G}_1 . The different colors distinguish the crosses corresponding to nodes of the three different communities. All crosses are above the 45° line.

3. THE EFFECTS OF NETWORK CHANGES

The structure of the MPA makes it easily adaptable to networks that change while the distributed computations are unfolding. New nodes and links may appear while other might disappear: if these changes happen without notice, the MPA cannot be restarted from the proper initial condition (3). For the sake of this discussion, we assume that the network changes only once, after the MPA has reached convergence on the initial network. We may interpret the dynamics that follows the change as an MPA running on the new network, but starting from with a different initialization. The result in Rossi and Frasca (2018) does not guarantee the convergence, because its proof uses a monotonicity property of the dynamics of $W^{i \rightarrow j}(t)$ that is only valid for the standard initialization (3). Moreover, it is not clear whether the MPA would carry some memory of the initial network.

In order to test these two facts, we construct a pair of small networks with partly similar topologies but very different harmonic influence profiles. The networks are $\mathcal{G}_3 = (I_3, E_3, C_3)$ and $\mathcal{G}_4 = (I_4, E_4, C_4)$, with $I_3 = I_4 = \{f, 1, \dots, 50\}$, while the edge sets E_3 and E_4 have the form of a “wheel” with additional connections. Both sets contain all the possible edges involving the field node, i.e. $\{f, i\}$ for $i \in \{1, \dots, 50\}$, and the cycle

$$\{\{1, 2\}, \{2, 3\}, \dots, \{49, 50\}, \{1, 50\}\}$$

connecting among all non-field nodes. Both sets also contain some extra edges of the form $\{i, j\}$ with $i \notin \{f, 1, 26\}$ and $j \neq f$: these are included with probability $p = 0.01$. Up to here, sets E_3 and E_4 are identical. To distinguish the networks, we include some additional edges $\{1, j\}$ in E_3 and some additional edges $\{26, j\}$ in E_4 : we pick these edges with probability $q = 0.25$. The matrix $C_3 \in \mathbb{R}_+^{I_3 \times I_3}$ has entries

$$\begin{cases} (C_3)_{if} = (C_3)_{fi} = 0.040 & \text{for every } i \in \{1, \dots, 50\} \\ (C_3)_{ij} = 1 & \text{if } i, j \neq f \text{ and } \{i, j\} \in E_3 \\ (C_3)_{ij} = 0 & \text{if } \{i, j\} \notin E_1 \end{cases}$$

The entries of C_4 are chosen similarly. The MPA starts on the network \mathcal{G}_3 and, after a sufficiently large time T ,

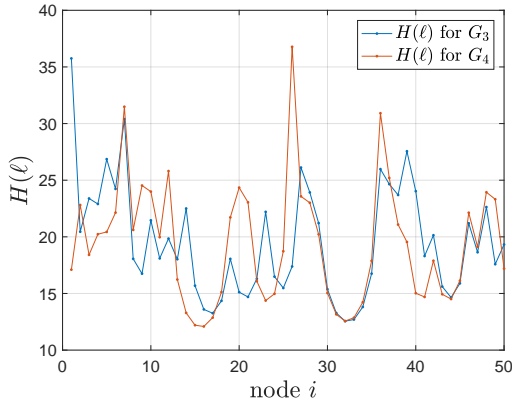


Fig. 6. The harmonic influence $H(\ell)$ of the nodes of \mathcal{G}_3 compared with the harmonic influence of the corresponding node in \mathcal{G}_4 .

continues on the network \mathcal{G}_4 . We use a bar to denote messages and estimates in this scenario. During the network change, the messages $\bar{W}^{i \rightarrow j}(T)$ and $\bar{H}^{i \rightarrow j}(T)$ corresponding to edges in $E_3 \cap E_4$ retain their values, while the messages corresponding to missing edges in $E_3 \setminus E_4$ are simply dropped. The messages corresponding to new edges $\{i, j\} \in E_4 \setminus E_3$ are initialized by

$$\bar{W}^{i \rightarrow j}(T) = 1, \bar{H}^{i \rightarrow j}(T) = 1 \text{ where } i \neq j.$$

We have repeated the simulation multiple times finding consistent results; we discuss one of the outcomes in what follows. The exact profiles of the harmonic influence are compared in Fig. 6. The most influential nodes are node 1 in \mathcal{G}_3 and node 26 in \mathcal{G}_4 ; their influences change significantly between the two graphs. Some other nodes hold very similar influences in \mathcal{G}_3 and \mathcal{G}_4 , e.g. nodes from 28 to 34.

We compare the convergence and estimates of the MPA on the changing scenario, with the convergence and estimate of the MPA started directly on \mathcal{G}_4 . The MPA on the changing network converges and requires less additional iterations (after T) to converge than the MPA on \mathcal{G}_4 : see Fig. 7. The convergence values $\bar{W}^{i \rightarrow j}(\infty)$ and $\bar{H}^\ell(\infty)$ on the changing network coincide exactly with those obtained directly on \mathcal{G}_4 . This result led us to conclude that the convergence values only depend on the final topology.

In conclusion, the MPA appears to be able to adapt smoothly to unforeseen changes in the network topology. Mathematically, we conjecture that the MPA has a unique equilibrium and converges to it from initial conditions that are more general than the standard ones in (3). While giving a full proof of this conjecture remains an open problem, we have so far verified that the equilibrium $\bar{W}^{i \rightarrow j}(\infty)$ is locally asymptotically stable.

REFERENCES

Acemoglu, D., Como, G., Fagnani, F., and Ozdaglar, A. (2013). Opinion fluctuations and disagreement in social networks. *Math. of Operations Research*, 38(1), 1–27.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.

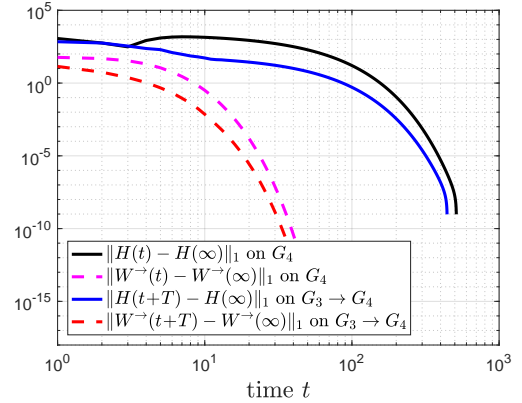


Fig. 7. The convergence of the MPA on the changing network scenario (after the change of topology) compared with the convergence over the network \mathcal{G}_4 . After the change of topology the MPA requires less iterations that a complete restart.

Fitch, K. and Leonard, N.E. (2016). Joint centrality distinguishes optimal leaders in noisy networks. *IEEE Trans. on Control of Network Systems*, 3(4), 366–378.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3), 75 – 174.

Leskovec, J. and McAuley, J.J. (2012). Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems 25*, 539–547.

Lin, F., Fardad, M., and Jovanovic, M.R. (2014). Algorithms for leader selection in stochastically forced consensus networks. *IEEE Transactions on Automatic Control*, 59(7), 1789–1802.

Rossi, W.S. and Frasca, P. (2018). On the convergence of message passing computation of harmonic influence in social networks. *IEEE Transactions on Network Science and Engineering*, PP(99), 1–1. doi: 10.1109/TNSE.2018.2792401.

Rossi, W.S. and Frasca, P. (2016). An index for the “local” influence in social networks. In *European Control Conference*, 525–530.

Rossi, W. and Frasca, P. (2017). Mean-field analysis of the convergence time of message-passing computation of harmonic influence in social networks. *IFAC-PapersOnLine*, 50(1), 2409–2414. 20th IFAC World Congress.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3), 1059–1069.

Van Mieghem, P., Devriendt, K., and Cetinay, H. (2017). Pseudoinverse of the Laplacian and best spreader node in a network. *Phys. Rev. E*, 96, 032311.

Vassio, L., Fagnani, F., Frasca, P., and Ozdaglar, A. (2014). Message passing optimization of harmonic influence centrality. *IEEE Transactions on Control of Network Systems*, 1(1), 109–120.

Yildiz, E., Ozdaglar, A., Acemoglu, D., Saberi, A., and Scaglione, A. (2013). Binary opinion dynamics with stubborn agents. *ACM Transactions on Economics and Computation*, 1(4), 1–30.