Harmonic influence in large-scale networks

Daron Acemoglu[*], Giacomo Como[†], Fabio Fagnani[‡] and Asuman Ozdaglar[§]

1. INTRODUCTION

Harmonic influence has been recently introduced as a measure of the relative influence of two nodes in a network that naturally emerges in models of opinion dynamics and social influence [1]. Given two nodes s_0 and s_1 in a connected network, that are assigned values $x_{s_0} = 0$ and $x_{s_1} = 1$, respectively, the harmonic influence vector x measures the relative influence of s_1 with respect to s_0 on the different nodes in the network. It is characterized by the property that the harmonic influence value x_v in any node $v \neq s_0, s_1$ coincides with the weighted average of the values of its neighbors. In other words, the harmonic influence vector is the solution of the Laplace equation on the network with boundary conditions on s_0 and s_1.

Harmonic influence can be given interpretations both in terms of random walks and electrical networks. More precisely, the value x_v coincides with the probability that a random walk on the network started in node v hits node s_1 before node s_0; on the other hand, x_v coincides with the voltage of node v in an electrical network where links' weights correspond to conductances and the voltages in s_0 and s_1 are fixed to the values 0 and 1, respectively. In fact, the connection to electrical networks has been exploited in the context of optimal placement of agents in a network with the purpose of swaying the average harmonic influence value. [6]

In [1], sufficient conditions for the harmonic influence vector to be almost constant throughout a large-scale network (a phenomenon referred to as *homogeneous influence*) were investigated. It was shown that harmonic influence is homogeneous in *highly fluid* networks, characterized by the property that the product between the mixing time of the associated stochastic matrix and the relative degree of the nodes s_0 and s_1 be vanishing in the large size limit.

In this work, we first study conditions under which harmonic influence *polarizes* in a large-scale network. Here, polarization refers to the existence of a cut in the network such that most of the nodes on the one side of it have harmonic influence value close to 0, and most of the nodes on the other side have value close to 1. In particular, we prove that, when the total size of the links between the two sides of a cut is negligible with respect to the degrees of s_0 and s_1, then the harmonic influence vector polarizes across this cut.

Then, we consider random interconnections between two highly fluid networks, one containing node s_0 and the other one containing node s_1 and prove the existence of a phase-transition. When the expected value of the total weight of the links interconnecting the two networks is negligible with respect to weight of s_0 and s_1, then harmonic influence polarizes across this cut. Conversely, when the weights of the interconnecting links are sufficiently concentrated around their expected value and their total expected value is much larger than the degree of s_0 and s_1, then harmonic influence is homogeneous.

Proofs, that are omitted here due to space limitations, rely on techniques from electrical networks and random walks theory [3, 2, 5].

2. HARMONIC INFLUENCE

Let a network be modeled as a connected undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, Q)$, where $\mathcal{V} = \{1, \dots, n\}$ is the set of nodes, $\mathcal{E}$ is the set of links, and $Q \in \mathbb{R}_+^{n\times n}$ is a symmetricmatrix with zero diagonal and such that $Q_{uv} > 0$ if and only if $\{u, v\} \in \mathcal{E}$. Let[1]

$$q := Q\mathbb{1}\,, \qquad \rho := \mathbb{1}'q\,, \qquad \chi := \frac{n}{\rho}\min_v q_v\,,$$

[1]Throughout, $\mathbb{1}$ stands for the all-one vector and $'$ for transpose.

[*]Department of Economics, MIT, USA. Email: daron@mit.edu.

[†]Automatic Control, Lund University, Sweden. Email: giacomo.como@control.lth.se.

[‡]Department of Mathematical Sciences, Politecnico di Torino, Italy. Email: fabio.fagnani@polito.it.

[§]Departement of EECS and LIDS, MIT, USA. Email: asuman@mit.edu.

Copyright is held by author/owner(s).

be the degree vector, the total degree, and the ratio between minimum and average degree. Let also

$$P = \mathrm{diag}\,(q)^{-1}Q\,, \qquad \pi := \rho^{-1}q\,;$$

notice that P is an irreducible stochastic matrix, π is its stationary probability distribution and P is reversible, i.e., $\pi_u P_{uv} = \pi_v P_{vu}$. We refer to a discrete-time Markov chain with transition probability matrix $\overline{P} = \frac{1}{2}(I + P)$ as the (lazy) random walk on $\mathcal{G}$.

Fix two nodes $s_0 \neq s_1 \in \mathcal{V}$, to be called *stubborn nodes*. We are interested in the unique vector $x \in \mathbb{R}^n$ solving the following linear system of equations

$$\begin{array}{ll} \sum_v Q_{uv}(x_v - x_u) = 0\,, & u \in \mathcal{V} \setminus \{s_0, s_1\}\,, \\ x_{s_i} = i\,, & i \in \{0,1\}\,. \end{array} \tag{1}$$

We will refer to x as the *harmonic influence* vector: its v-th entry, x_v, measures the relative influence on node v exerted by node s_1 with respect to that exerted by node s_0. Observe that (1) is equivalent to

$$\begin{array}{ll} x_u = \sum_v P_{uv} x_v\,, & u \in \mathcal{V} \setminus \{s_0, s_1\}\,, \\ x_{s_i} = i\,, & i \in \{0,1\}\,. \end{array} \tag{2}$$

that is, the value of every node other than s_0 and s_1 is the weighted average of its neighbor nodes' values.

Existence and uniqueness of the harmonic influence vector are standard facts (see, e.g., [5, Proposition 9.1]). The weighted mean of vector x's entries

$$\overline{x} := \pi' x = \frac{1}{\rho}\sum_v q_v x_v$$

will be called the *average harmonic influence* value. Note that the case when there are multiple nodes whose value is fixed to either 0 or 1 can be treated by just considering the network where all the nodes with value 0 are collapsed in a single node s_0, and all the nodes with value 1 are collapsed in a single node s_1. As proven in [1], x can be thought of as the vector of expected stationary opinions of a stochastic opinion dynamics with gossip or voter opinion updates.

In the following, we investigate properties of the harmonic influence vector x, with particular focus on large-scale networks. These are modeled as sequences of networks of increasing size n, and we will concentrate on asymptotic behaviors as n grows large. When doing so, limits have to be intended always as $n \to +\infty$, unless specified otherwise. We will also use the Landau notation, writing $a = o(b)$ for $\lim a/b = 0$, $a = O(b)$ meaning that $a \leq Kb$ for some positive constant K independent of n, and $a \asymp b$ for $a = O(b)$ and $b = O(a)$. Finally, we will say that some property holds 'with high probability' if the probability that the property holds converges to 1 as n grows large. When considering large-scale networks, we always assume that

$$\liminf \chi > 0\,, \tag{3}$$

i.e., that the ratio between the minimum and average degree remains bounded away from 0 as n grows large, a property which is satisfied by most of the large-scale networks considered in the literature. Note that this is very different from requiring the ratio between the maximum and average degree to remain bounded, a property which is not satisfied by many large-scale networks.

3. HOMOGENEOUS INFLUENCE VS POLARIZATION

In [1], sufficient conditions for the harmonic influence to be homogeneous were derived. Precisely, *homogeneous influence* is meant to be the property that

$$\frac{1}{n}|\{v : |x_v - \overline{x}| \geq \varepsilon\}| \to 0\,, \qquad \forall \varepsilon > 0\,, \tag{4}$$

i.e., that $x_v = \overline{x} + o(1)$ for all but a vanishing fraction of nodes. Such conditions can be formulated in terms of the *mixing time* of the matrix P, defined as

$$\tau_{\mathrm{mix}} := \inf\left\{t \geq 0 : \max_{u,v}\sum_w \left|(\overline{P}^t)_{uw} - (\overline{P}^t)_{vw}\right| \leq \frac{1}{2e}\right\}.$$

The mixing time measures the time required to the random walk on $\mathcal{G}$ to get close to stationarity. As is well known [2, 5], τ_{mix} can be estimated in terms of the network conductance, i.e., its smallest bottleneck ratio. In particular, such estimates imply that P is fast mixing, i.e., τ_{mix} grows at most (poly)logarithmically in n, when the conductance is either bounded away from 0 or decreases at most polylogarithmically in n. This is known to be the case in many random large-scale networks of interest such as Erdos-Renyi graphs in the connected regime, configuration models, preferential attachment graphs, and small worlds. [4]

Theorem 4 in [1], implies that

$$\frac{1}{n}|\{v : |x_v - \overline{x}| \geq \varepsilon\}| \leq \frac{1}{\chi\varepsilon}\theta(\tau_{\mathrm{mix}}(\pi_{s_0} + \pi_{s_1}))\,, \tag{5}$$

where $\theta(y) := y \log e^2/y$, for all $y > 0$. Inequality (5) combined with our standing assumption (3), implies that in *highly fluid* networks, characterized by the property that

$$\tau_{\mathrm{mix}}(\pi_{s_0} + \pi_{s_1}) = \tau_{\mathrm{mix}}\rho^{-1}(q_{s_0} + q_{s_1}) \to 0\,, \tag{6}$$

influence is homogeneous. Several examples of highly fluid networks are reported in [1, Sect.6] including all the aforementioned cases of fast mixing large-scale random networks when s_0 and s_1 are obtained by merging multiple nodes with total degree $O(n^{1-\varepsilon})$.

We now shift focus towards studying polarization in large-scale networks. Consider a *relative cut* in the network $\mathcal{G}$ separating s_0 from s_1. For $i = 0, 1$, let $\mathcal{V}_i$ be the part of the node set $\mathcal{V}$ containing node s_i, $\partial_i := \{u \in \mathcal{V}_i : \{u, v\} \in \mathcal{E}$ for some $v \in \mathcal{V}_{i-i}\}$ be the internal boundary of $\mathcal{V}_i$, and $\mathcal{G}_i$ be the network obtained from $\mathcal{G}$ by collapsing $\mathcal{V}_{1-i}$ into a single node

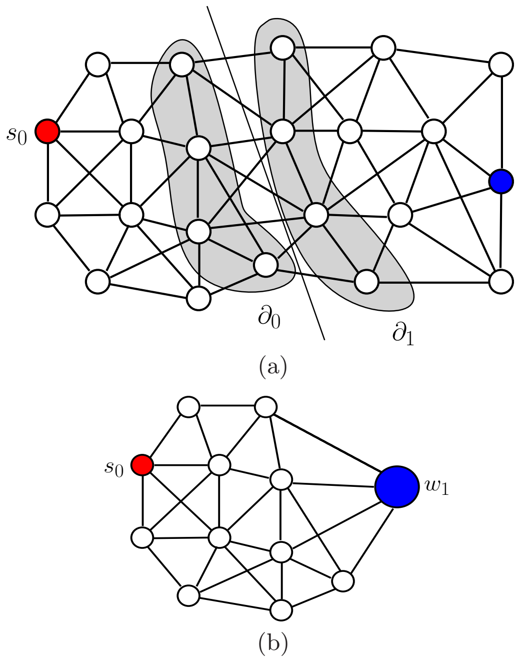

Figure 1: In (a) a cut in $\mathcal{G}$ separating s_0 from s_1, where the internal boundaries ∂_0 and ∂_1 are shaded in gray. In (b), the subnetwork $\mathcal{G}_0$.

to be denoted by w_{i-1}, so that the node set of $\mathcal{G}_i$ is $\mathcal{V}_i \cup \{w_{i-1}\}$ (see Figure 1). In the special case when $\mathcal{V}_i = \{s_i\}$ and $\mathcal{V}_{1-i} = \mathcal{V} \setminus \{s_i\}$, one gets that $\mathcal{G}_i$ is a simple network with node set $\{s_i, w_{1-i}\}$ and one link of weight q_{s_i}, while $\mathcal{G}_{1-i}$ coincides with the original network $\mathcal{G}$. More in general, we shall assume that, $\mathcal{G}^i$ is connected for $i = 0, 1$ and let $n_i := |\mathcal{V}_i|$, π^i, and τ^i_{mix} be the corresponding size, invariant probability distribution and mixing time, respectively. Let also

$$\alpha := \rho^{-1} \sum_{u \in \mathcal{V}_0} \sum_{v \in \mathcal{V}_1} Q_{uv}$$

be the relative weight of the cut.

We shall say that the i-th subnetwork *polarizes* if

$$\frac{1}{n_i} |\{v \in \mathcal{V}_i : |x_v - i| > \varepsilon\}| \to 0 \,, \qquad (7)$$

i.e., if the harmonic influence value of all but a vanishing fraction of nodes on s_i's side of the cut converges to i. One may conjecture that the i-th subnetwork be highly fluid and $\alpha/q_{s_i} \to 0$ would be sufficient conditions for the i-th subnetwork to polarize on s_i. While such a conjecture can be disproved as such, we are going to formulate a refined version of it that can be proven to hold true. Let us define the *escape probability* from a node $v \in \mathcal{V}_i \cup w_{1-i}$ as

$$\zeta^i_v := \sup_{k \geq 0} \frac{\mathbb{P}^i_v(T^+_v > k\tau^i_{\text{mix}}) - 2e^{-k}}{1 + k\tau^i_{\text{mix}}\, \pi_v} \,. \qquad (8)$$

In the above, the symbol $\mathbb{P}^i_v$ refers to the probability for a random walk on $\mathcal{G}_i$ started from node v, and T^+_v stands for the return time of such random walk to node v. It is not hard to verify that $0 \leq \zeta^i_v \leq 1$ (the second inequality is immediate, for the first one it is sufficient to consider $k \to \infty$). The reason for the terminology comes from the fact that $\mathbb{P}^i_v(T^+_v > k\tau^i_{\text{mix}})$ is the probability that the random walk spends more than $k\tau^i_{\text{mix}}$ time steps before returning to its starting node v: this term, which is clearly non increasing in k, is then combined with the increasing term $-2e^{-k}$, normalized by the factor $(1 + k\tau^i_{\text{mix}}\, \pi_v)$, and optimized over choices of $k \geq 0$. While the specific form of the right-hand side of (8) results from the technical details of the proofs that are omitted here, it is possible to understand when it occurs that ζ^i_v is strictly larger than 0. This is the case when one can find some positive k such that $\mathbb{P}^i_v(T^+_v > k\tau^i_{\text{mix}}) > 2^{-k}$. For large-scale networks, one has that

$$\liminf \zeta^i_v > 0 \qquad (9)$$

if $\tau^i_{\text{mix}}\, \pi^i_v \to 0$ and $\liminf \mathbb{P}^i_v(T^+_v > k\tau^i_{\text{mix}}) > 0$ for some k which grows large with the network size n. From now on, we shall refer to (9) as the property of *positive escape probability* from node s_i in a large-scale network. It can be proven that the random large-scale networks mentioned before (i.e, connected Erdos-Renyi, configuration models, preferential attachment, and small worlds) have finite escape probability from nodes obtained by merging random nodes with total degree $O(n^{1-\varepsilon})$.

We are now ready to formulate the first main result of this contribution.

Theorem 1. *Consider a large scale network $\mathcal{G}$ and a cut separating s_0 from s_1. Assume that for $i \in \{0, 1\}$, the subnetwork $\mathcal{G}_i$ is highly fluid, i.e, $\tau^i_{\text{mix}}\, \pi_{s_i} = \tau^i_{\text{mix}}\, q_{s_i}/\rho \to 0$, and have positive escape probability from node s_i, i.e., $\liminf \zeta^i_{s_i} > 0$. If $\alpha/\pi_{s_i} = \alpha\rho/q_{s_i} \to 0$, then $\mathcal{G}_i$ polarizes, as for* (7).

The intuition behind this result is the following: the condition $\alpha/\pi_{s_i} \to 0$ implies that the total weight of links across the cut is negligible with respect to the degree of node s_i, so that the influence of node s_i dominates the one of all the nodes on the other side of the cut (including s_{1-i}). The condition $\tau^i_{\text{mix}}\, \pi_{s_i} \to 0$ implies that influence is homogeneous in $\mathcal{V}_i$, similarly to Theorem 4 in [1]. Finally, the assumption of positive escape probability guarantees that dominance of the influence of s_i on s_{1-i} is not limited to the immediate neighborhood of s_i, but can spread through the network and affect most of $\mathcal{V}_i$. As mentioned, the result does not hold true without this last assumption.

Theorem 1 should be contrasted with Theorem 4 in [1]. While the latter states that influence is homogeneous in highly fluid networks characterized by the absence of small bottlenecks, when applied to both $i = 0$ and $i = 1$, the former states that when the cut between two internally highly fluid subnetworks has a weight negligible with respect to the stubborn nodes, then the two subnetworks polarize each on the value of the corresponding stubborn node.

In fact, one can obtain more direct converse results to Theorem 1. A natural conjecture is

Conjecture 1. *Consider a large scale network $\mathcal{G}$ and a cut separating s_0 from s_1. For $i \in \{0,1\}$, let the subnetwork $\mathcal{G}_i$ be such that $\tau^i_{\rm mix}\pi_{w_{i-1}} \to 0$ and the escape probability from node w_{i-1} is finite, i.e., $\liminf \zeta^i_{w_{i-1}} > 0$. Then, harmonic influence is homogeneous provided that and $\pi_{s_i}/\alpha \to 0$, for $i = 0,1$.*

We have been able to prove the following two weaker versions of Conjecture 1. The first one involves a relaxation of the notion of homogeneous influence. Let

$$z^*_0 := \max_{v_0} x_{v_0}\,, \qquad z^*_1 := \min_{v_1} x_{v_1}$$

where, while intended to run over $\mathcal{V}_0$ and $\mathcal{V}_1$, the maximation/minimization indices can actually be restricted to ∂_0 and ∂_1, respectively. We use the term *weakly homogeneous influence* with the meaning that

$$\frac{1}{n}\,|\{v:\, z^*_1 - \varepsilon \le x_v \le z^*_0 + \varepsilon\}| \to 1\,, \quad \forall \varepsilon > 0\,. \quad (10)$$

Weakly homogeneous influence implies that $z^*_1 \le z^*_0 + o(1)$. When $|\partial_1| = 1$ or $|\partial_0| = 1$, the maximum principle implies that $z^*_0 \le z^*_1$, so that (10) implies (4), i.e., weakly homogeneous influence is the same as homogeneous influence. In general, this may not be the case, and what is missing from (10) in order to get (4) is an inequality of the form $z^*_0 \le z^*_1 + o(1)$.

Then, the following result follows from Theorem 1.

Proposition 1. *Consider a large scale network $\mathcal{G}$ and a cut separating s_0 from s_1. For $i \in \{0,1\}$, let the escape probability from node w_{i-1} be positive in the subnetwork $\mathcal{G}_i$, i.e., $\liminf \zeta^i_{w_{i-1}} > 0$. If $\tau^i_{\rm mix}\pi_{w_{i-1}} \to 0$ and $\pi_{s_i}/\alpha \to 0$, for $i = 0,1$, then harmonic influence is weakly homogenous.*

On the other hand, one can strengthen the assumptions of Conjecture 1 instead of weakening its conclusion. For $i = 0,1$, let $\chi_i := \max_{u\in\mathcal{V}_i}(\alpha q_u)^{-1}\sum_{v\in\mathcal{V}_{1-i}} Q_{uv}$ measure the maximum ratio between the relative weight of the connections of a node to the other side of the cut, and the relative weight of the cut. We have

Proposition 2. *If $\alpha\tau^i_{\rm mix} \to 0$, $\pi_{s_i}/\alpha \to 0$, and*

$$\limsup \chi_i < +\infty\,, \quad (11)$$

for $i = 0,1$, then influence is homogeneous.

It should be noted that the statement above allows one to prove the original notion of homogeneous influence (4) as opposed to the potentially weaker one (10). Moreover, it does not require any assumption of positive escape probability. On the other hand, it requires the additional assumption (11) that basically amounts to that the boundaries ∂_0 and ∂_1 be a non-vanishing fraction of the respective node sets.

4. PHASE TRANSITIONS FROM POLARIZATION TO HOMOGENEITY

Based on the results in the previous session, it is possible to analyze large-scale networks where, at the change of a parameter, the network transitions from a condition of complete polarization to one of homogeneous influence. This can be done as follows. For $i = 0,1$, let $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i, Q^i)$ be two rapidly mixing networks of comparable size, and let $s_i \in \mathcal{V}_i$ be two stubborn nodes (possibly obtained by merging together multiple nodes) such that $q_{s_0} + q_{s_1} = O(n^{1-\varepsilon})$ and both have positive escape probability $\liminf \zeta^i_{s_i} > 0$. Then, consider a network $\mathcal{G}$ obtained by interconnecting $\mathcal{G}_0$ and $\mathcal{G}_1$ as follows: between any pair $\{v_0, v_1\}$ with $v_0 \in \mathcal{V}_0 \setminus \{s_0\}$ and $v_1 \in \mathcal{V}_1 \setminus \{s_1\}$ there is a weight-β link independently with probability γ.

Then, Theorem 1 implies that

(a) if, $n^2\beta\gamma/q_{s_i} \to 0$, then, with high probability, most of the i-th subnetwork polarizes on the value of its stubborn node.

On the other hand, Proposition 1 implies that

(b1) if $n^2\beta\gamma/(q_{s_0}+q_{s_1}) \to +\infty$, and $\beta\gamma = O(1/n^{1+\varepsilon})$ for some $\varepsilon > 0$, then, with high probability, influence is weakly homogeneous.

This can be strengthened using Proposition 2 to

(b2) if $n^2\beta\gamma/(q_{s_0} + q_{s_1}) \to +\infty$, $\liminf \gamma > 0$, and $\beta = O(1/n^{1+\varepsilon})$, then, with high probability, influence is homogeneous.

Hence, if we consider the parameter $\beta\gamma/q_s$, then $1/n^2$ is a threshold function for polarization vs (weakly) homogeneous influence property.

5. CONCLUSION

In this work, we have studied harmonic influence in large-scale networks. We have characterized sufficient conditions for the network to be polarized and investigated the existence of a phase transition between homogeneous influence and polarization.

6. REFERENCES

[1] D. Acemoglu, G. Como, F. Fagnani, and A. Ozdaglar, *Opinion fluctuations and disagreement in social networks*, Mathematics of Operation Research **38** (2013), no. 1, 1–27.

[2] D. Aldous and J. Fill, *Reversible Markov chains and random walks on graphs*, Monograph, 2002.

[3] P. Doyle and J. Snell, *Random walks and electric networks*, Math. Association of America, 1984.

[4] R. Durrett, *Random graph dynamics*, Cambridge University Press, 2006.

[5] D.A. Levin, Y. Peres, and E.L. Wilmer, *Markov chains and mixing times*, AMS, 2010.

[6] L. Vassio, F. Fagnani, P. Frasca, and A. Ozdaglar, *Message passing optimization of harmonic influence centrality*, IEEE Trans. Control of Network Systems (2014).