Triadic Closure Alleviates Network Segregation

Rediet Abebe[*]
Harvard University
rabebe@fas.harvard.edu

Nicole Immorlica
Microsoft Research
nicimm@microsoft.com

Jon Kleinberg [†]
Cornell University
kleinber@cs.cornell.edu

Brendan Lucier
Microsoft Research
brlucier@microsoft.com

Abstract

Homophily – the tendency for individuals to form social ties to others who are similar to themselves – is one of the most robust sociological principles, leading to patterns of linkage in social networks that segregate people along many different demographic dimensions. This phenomena, in turn, can result in inequalities in access to information and opportunities by members of different demographic groups. As we consider potential interventions that might alleviate the effects of network segregation, we face the challenge that homophily constitutes a pervasive and organic force that is difficult to push back against. The design of interventions can therefore benefit from identifying counterbalancing natural processes in the network that might be harnessed to work in opposition to segregation. In this work, we examine several fundamental network formation models to show that *triadic closure* is one such process. Our analyses show the power for triadic closure to reduce network segregation at the level of graph structure. We believe that the insights in this work have qualitative implications for the design of network interventions in settings such as online platforms and college dorm assignments, where the designer has a vested interest in mitigating network segregation.

1 Introduction

Across societies, the tendency for individuals to form social ties with others with whom they share similarities is a robust and pervasive social phenomena impacting the formation of social networks [26, 27, 29, 30]. This phenomena, known as *homophily*, can result in segregation; and since networks play a key role in the diffusion of information, opportunities, and resources, it can lead to inequalities across members of different communities; empirical and theoretical work has shown that network structures can influence individuals' ability to obtain accurate and relevant information, garner social support, improve their labor market outcomes, among many other impacts [4, 7, 8, 18, 21, 35, 31].

Understanding how network formation processes are impacted by homophily is important for predicting and improving societal welfare in various domains. However, since homophily is a potent and organic force, it is challenging to push back against its negative consequences without identifying other social processes that may already be working against segregation. By pinpointing such network forces and understanding their interactions with homophily, we can then harness their impact to improve network diversity and welfare of individuals across the network.

[*]RA was supported a Facebook Scholarship, a Simons Investigator grant, and a MacArthur Foundation grant.
[†]JK was supported in part by a Simons Investigator Award, a Vannevar Bush Faculty Fellowship, a MURI grant, and a MacArthur Foundation grant.

Joint Workshop on AI for Social Good (NeurIPS 2019), Vancouver, Canada.

In this work, we show that *triadic closure* – the process by which individuals are more likely to form social ties with other individuals with whom they already share ties – may be one such force [16, 25, 32]. Through analyses of different fundamental and popular network formation models, we show that triadic closure has the tendency to reduce network segregation. That is, even though biases towards homophily in link formation tend to push a network towards segregation, triadic closure has the effect of exposing people to dissimilar individuals, thereby resulting in more integrated networks. We find it striking that such a tension should exist between two such well-studied social processes as homophily and triadic closure, and we believe it points to the possibility of broader phenomena that we begin exploring in this paper.

The tension between these two forces is also partly counter-intuitive: triadic closure produces links between people with common friends, and we might therefore suspect that these new links reinforce the underlying similarity that the friends share; but the longer-range nature of triadic closure, connecting people separated by multiple steps in the network, in fact has the aggregate effect of exposing nodes to others who are dissimilar in the models that we consider. Indeed much work, including recent work, has asserted that triadic closure can result in social segregation at the macroscale [34, 37].

In this work, we explore this relationship between these phenomena by analyzing the effect of triadic closure on the proportion of links between dissimilar agents on different static and dynamic network formation models. The models we consider are characterized by the presence of heterogeneous nodes and a two phase link-formation process where agents first form a fixed set of links followed by a triadic closure phase were they form an additional set of links through a friend-of-friend search.

2 Triadic Closure and Stochastic Block Models

We begin with a simple static model based on stochastic block models (SBM). We suppose that there are n nodes corresponding to individuals, where each node can be one of k types. We assume that there are an equal number of nodes of each type. Note, we assume that all edges formed in this model and the next one are undirected.

In an SBM with k types, there is a probability p of an edge forming between two nodes of the same type (this is the *in-block probability*) and a probability q of an edge forming between two nodes of different types (this is the *out-of-block probability*). All edge formations are mutually independent. We first run an SBM to obtain a network G. We then consider the effect of triadic closure on G. In order to do so, we must first study wedges in G.

A *wedge* W in a graph G is an induced subgraph on nodes $\{v_1, v_2, v_3\}$ such that there exists edges (v_1, v_2) and (v_2, v_3), but there exists no edge between nodes v_1 and v_3. We denote this wedge W using (v_1, v_2, v_3).

Definition 1. We say $W = (v_1, v_2, v_3)$ is a *monochromatic wedge* if v_1 and v_3 share a type. Otherwise, we say it is a *bichromatic wedge*.

Definition 2. Given a set of nodes $T = \{v_1, v_2, v_3\}$, we say that T is a *monochromatic triplet* if all three nodes share a type and we say that it is a *bichromatic triplet* otherwise.

Monochromatic and bichromatic wedges correspond to wedges where the missing edges are monochromatic and bichromatic, respectively. Note, further, that a monochromatic wedge can result from either a monochromatic or a bichromatic triplet, whereas a monochromatic triplet can only yield a monochromatic wedge. An object of interest in this section is the fraction of bichromatic wedges, which we denote by $w(G)$. We can estimate this value using the following result, presented here for the case where $k = 2$:

Lemma 3. *Given a network G resulting from an SBM, the expected number of monochromatic wedges is*

$$3 \cdot 2 \cdot \binom{n/2}{3} \cdot p^2(1-p) + n \cdot \binom{n/2}{2} \cdot q^2(1-p), \tag{1}$$

while that of bichromatic wedges is

$$2 \cdot n \cdot \binom{n/2}{2} \cdot pq(1-q). \tag{2}$$

This follows a straightforward counting argument. The proof of this lemma as well as the subsequent theorem are presented for the general case of $k \geq 2$ in the appendix.

We then consider the effect of triadic closure on network integration. Specifically, we measure network integration using the proportion of bichromatic edges: i.e., the fraction of edges in G between nodes of dissimilar types, denoted by $f(G)$. We say that triadic closure *increases network integration* if, by closing a single randomly selected wedge, we increase the fraction of bichromatic edges. We can likewise define decreasing network integration and preserving network integration.[3]

Theorem 4. *Given an SBM, there is a sufficiently large n such that triadic closure increases network integration if $p > q$ and it preserves network integration if $p = q$.*

3 A Dynamic Network Formation Model

We now consider a dynamic process of network formation, based on a natural variation of a popular growing graph model by Jackson and Rogers [22]. In this Jackson-Rogers model, homogeneous nodes arrive sequentially and form links first through a *random meeting phase*, where a new node selects random nodes to link with, and next through a *network search phase*, where this node expands its set of neighbors through linking with its friends-of-friends.

We adapt this model to account for heterogeneous nodes. As in the previous section, we assume that each node can be one of two types and that there are an equal number of nodes of each type. The model proceeds as follows: nodes arrive to the network consecutively. When node v arrives

I it selects N_S neighbors uniformly at random from nodes of the same type and N_D neighbors uniformly at random from nodes of a different type, then

II it selects N_F additional neighbors according to the following biased friend-of-friend search process: suppose $\alpha \in (0, 1]$. Let $F_S(v)$ be the set of same-type neighbors that v selected in the first phase and $F_S^2(v) = \{w : u \in F_S(v),\ (u, w) \in G(t-1)\}$ be the neighbors of v's same-type neighbors. (We can similarly define $F_D(v)$ and $F_D^2(v)$.) Then v selects αN_F nodes uniformly at random among $F_S^2(v)$ and $(1-\alpha)N_F$ nodes uniformly at random among $F_D^2(v)$ and creates links.

In this above model, time moves in discrete steps $t = \{0, 1, 2, \ldots\}$, and $G(t)$ denotes the corresponding graph at time t. We let $N := N_S + N_D + N_F$ and so the number of edges in $G(t)$, which we denote by $m(t)$, is Nt. We assume that N_S, N_D, N_F are all fixed constant values.

We assume here that nodes show *type-bias*. i.e., that $N_S > N_D$ and that $\alpha > \frac{N_S}{N_D+N_S}$.

We denote the equilibrium state network, that is $G(t)$ as $t \to \infty$, by G. To ensure that this process is well defined, we let $G(0)$ a stochastic block model on N nodes and we let each node select $\frac{N_F}{N_D+N_S}$ nodes to form links with uniformly at random. We assign each node a random type.

We are interested in the network integration, as measured by the fraction of bichromatic edges, in equilibrium. We denote by $f(t)$ the fraction of bichromatic edges at time t. Here, we want to show that $f(t)$ converges and that we can fully characterize its value in equilibrium. Let $b(t)$ be the number of bichromatic edges at time t and recall that $m(t)$ is the number of all edges. Therefore, $f(t) = b(t)/m(t)$.

Given a node v, we let $f_v(t)$ be the fraction of v's neighbors that do not share a type with v. We can then define

$$f_R(t) = \frac{1}{|R|} \sum_{r \in R} f_r(t).$$

Note that this is an average value over all nodes of type R. We can similarly define $f_B(t)$.

Since this is a growing graph, the degree of nodes also evolves. Namely, as time proceeds, each node has more opportunities to form more links. We denote the degree of node v at time t by $\Delta_v(t)$ and

[3] Note that these definitions are local: they analyze the impact of a single triadic closure. Note, however, that for large n, this observation on network integration applies for the case where we can instead consider the effect of an unbounded but sublinear number of simultaneous triadic closures.

define

$$\Delta_R(t) = \frac{1}{|R|}\sum_{r\in R}\Delta_r(t).$$

That is, $\Delta_R(t)$ is the average degree of an R node at time t.

Without loss of generality, suppose a new node v of type R is added at time $t+1$. Then,

$$b(t+1) = b(t) + N_D + \alpha N_F f_R(t) + (1-\alpha)N_F(1 - f_B(t)) \tag{3}$$

Our analysis below involves a step in which we establish an identity (Identity 1, below) on the long-run fraction of bichromatic edges $f(t)$ using a heuristic argument that is supported by numerical simulation. The identity is as follows. Let f^* denote the limiting fraction of bichromatic edges of $G(t)$ as t goes to infinity. Then we have

$$\textit{(Identity 1)} \qquad f^* = \frac{N_D + (1-\alpha)N_F}{N_D + N_S + 2(1-\alpha)N_F}.$$

We note that $f(t)$ shows that the fraction of bichromatic links formed in the first phase can have a disproportionate impact on the network integration in equilibrium. For instance, for $\alpha = 1$, which corresponds to the instance where all friend-of-friend links are formed through same-type friends, $f(t) = \frac{N_D}{N_S+N_D}$, indicating that this value is *exclusively* determined by the first phase despite the assumption that $N_S + N_D$ is much less than N_F. The role of N_F in the fraction of bichromatic edges in equilibrium increases as α decreases. Further, for *any* $\alpha < 1$, $f(t)$ approaches $1/2$ as N_F grows large relative to N_S and N_D. So even a single off-type seed friend can have a significant down-stream impact, if triadic closure is not perfectly driven by same-type friends.

3.1 Triadic Closure and Network Integration.

As above, we say that triadic closure *increases network integration* if f^* goes up as N_F goes up, we say that it *decreases network integration* if f^* goes down as N_F goes down, and that it *preserves network integration* if f^* is unaffected by N_F. Recall that we assume that nodes show type-bias: i.e., $N_S > N_D$ and $\alpha > \frac{N_S}{N_D+N_S}$.

Theorem 5. *Assuming Identity 1, triadic closure increases network integration in our dynamic network formation model if $\alpha < 1$ if nodes show type-bias in the first phase and preserves network integration if nodes show no bias. Triadic closure preserves network integration for $\alpha = 1$.*

Proof. Select $N_F, N_{F'}$ such that $N_F < N_{F'}$. Denote by f^* and f'^* the fraction of bichromatic links in equilibrium state for N_F and $N_{F'}$, respectively. We are interested in the inequality $f^* \le f'^*$, which we can rewrite as

$$\frac{N_D + (1-\alpha)N_F}{N_D + N_S + 2(1-\alpha)N_F} \le \frac{N_D + (1-\alpha)N_{F'}}{N_D + N_S + 2(1-\alpha)N_{F'}}$$

We can simplify this above inequality to get

$$2N_D(N_{F'} - N_F) \le (N_S + N_D)(N_{F'} - N_F).$$

This holds with strict inequality if $N_D < N_S$ and with equality if $N_D = N_S$.

For the degenerate case where $\alpha = 1$, note f^* and $f^{'*}$ are both $\frac{N_D}{N_S+N_D}$. That is, the fraction of bichromatic edges is determined exclusively by the random meeting phase and triadic closure preserves network integration in this setting. □

Note that for the case where $N_S = N_D$, triadic closure preserves network integration even if nodes show type-bias in the second phase, by forming a disproportionate number of links through their same-type friends. Likewise, for the case where $\alpha = 1$, triadic closure preserves network integration regardless of the type-bias that nodes show in the first phase. Outside of these two degenerate cases, we note that triadic closure has the effect of increasing network integration.

4 Conclusion and Future Work

In this work, we consider the effect of triadic closure on network segregation in this work. Through analysis of different static and dynamic network formation models, we find that triadic closure has the effect of increasing network integration, indicating that it may be a process that counteracts homophily in network formation. We view the insights in this paper as potentially pointing to broader phenomena about these social processes, which can be studied both mathematically as well as analysis of real-world networks.

The results presented in this work open up questions related to other measurements of network health, such as network expansion and distribution of network centralities. Each of these points to challenging analytic questions. Empirically, it would also be interesting to shed light on what types of social and information networks tend to exhibit a stronger relationship between triadic closure and homophily.

In ongoing work, we find that the insights above can help inform network interventions to minimize segregation. Namely, in many scenarios, an authority that has a vested interest in the outcome of a network process can attempt to nudge the network towards a more integrated state. For example, on-line networking platforms recommend links to create a warm start for new users, as well as to fill out established networks. We find evidence that minor and local interventions on homophily acting in the early stages of a network formation process can be amplified through triadic closure, resulting in an indirect yet substantial shift on the network integration. Through ongoing work, we are exploring both the theoretical problem of finding optimal network interventions and empirically analyzing their effect on network integration.

References

[1] Lada A. Adamic and Natalie Glance. The political blogosphere and the 2004 u.s. election: Divided they blog, 2005.

[2] Kristen M Altenburger and Johan Ugander. Monophily in social networks introduces similarity among friends-of-friends. *Nature human behaviour*, 2(4):284, 2018.

[3] Chen Avin, Barbara Keller, Zvi Lotker, Claire Mathieu, David Peleg, and Yvonne-Anne Pignolet. Homophily and the glass ceiling effect in social networks, 2015.

[4] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144), 2013.

[5] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. Equity of attention: Amortizing individual fairness in rankings. *arXiv preprint arXiv:1805.01788*, 2018.

[6] Yann Bramoullé, Sergio Currarini, Matthew O Jackson, Paolo Pin, and Brian W Rogers. Homophily and long-run integration in social networks. *Journal of Economic Theory*, 147(5), 2012.

[7] Antoni Calvo-Armengol and Matthew O Jackson. The effects of social networks on employment and inequality. *American Economic Review*, 94(3), 2004.

[8] Antoni Calvo-Armengol, Eleonora Patacchini, and Yves Zenou. Peer effects and social networks in education. *The Review of Economic Studies*, 76(4), 2009.

[9] Sergio Currarini, Matthew O Jackson, and Paolo Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

[10] Krishna Dasaratha. Distributions of centrality on networks. *arXiv preprint arXiv:1709.10402*, 2017.

[11] Paul DiMaggio and Filiz Garip. How network externalities can exacerbate intergroup inequality. *American Journal of Sociology*, 116(6):1887–1933, 2011.

[12] Yuxiao Dong, Reid A Johnson, Jian Xu, and Nitesh V Chawla. Structural diversity and homophily: A study across more than one hundred big networks, 2017.

[13] Nathan Eagle, Michael Macy, and Rob Claxton. Network diversity and economic development. *Science*, 328(5981):1029–1031, 2010.

[14] Michael D Ekstrand and Martijn C Willemsen. Behaviorism is not enough: better recommendations through listening to users. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 221–224. ACM, 2016.

[15] Jacob K Goeree, Arno Riedl, and Aljaž Ule. In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior*, 67(2), 2009.

[16] Mark S Granovetter. The strength of weak ties. In *Social networks*, pages 347–367. Elsevier, 1977.

[17] Ido Guy. Social recommender systems. In *Recommender Systems Handbook*, pages 511–543. Springer, 2015.

[18] Anikó Hannák, Claudia Wagner, David Garcia, Alan Mislove, Markus Strohmaier, and Christo Wilson. Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr., 2017.

[19] Adam Douglas Henry, Paweł Prałat, and Cun-Quan Zhang. Emergence of segregation in evolving social networks. *Proc. of the National Academy of Sciences*, 108(21), 2011.

[20] Jevan Hutson, Jessie G Taft, Solon Barocas, and Karen Levy. Debiasing desire: Addressing bias & discrimination on intimate platforms. *arXiv preprint arXiv:1809.01563*, 2018.

[21] Matthew O. Jackson, Tomas Rodriguez-Barraquer, and Xu Tan. Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, 102(5), May 2012.

[22] Matthew O. Jackson and Brian W. Rogers. Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915, June 2007.

[23] Kibae Kim and Jörn Altmann. Effect of homophily on network formation. *Comm. in Nonlinear Science and Numerical Simulation*, 44, 2017.

[24] Bart P Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. Recommender systems for self-actualization. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 11–14. ACM, 2016.

[25] Gueorgi Kossinets and Duncan J Watts. Empirical analysis of an evolving social network. *science*, 311(5757):88–90, 2006.

[26] Gueorgi Kossinets and Duncan J Watts. Origins of homophily in an evolving social network. *American journal of sociology*, 115(2):405–450, 2009.

[27] Paul F Lazarsfeld, Robert K Merton, et al. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66, 1954.

[28] Adalbert Mayer and Steven L Puller. The old boy (and girl) network: Social network formation on university campuses. *Journal of Public Economics*, 92(1-2), 2008.

[29] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1), 2001.

[30] Mark EJ Newman. Assortative mixing in networks. *Physical review letters*, 89(20):208701, 2002.

[31] Shirin Nilizadeh, Anne Groggel, Peter Lista, Srijita Das, Yong-Yeol Ahn, Apu Kapadia, and Fabio Rojas. Twitter's glass ceiling: The effect of perceived gender on online visibility., 2016.

[32] Anatol Rapoport. Spread of information through a population with socio-structural bias: I. assumption of transitivity. *The bulletin of mathematical biophysics*, 15(4):523–533, 1953.

[33] Tobias Schnabel, Paul N Bennett, Susan T Dumais, and Thorsten Joachims. Short-term satisfaction and long-term coverage: Understanding how users tolerate algorithmic exploration. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 513–521. ACM, 2018.

[34] Christoph Stadtfeld. The micro–macro link in social networks. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, pages 1–15, 2015.

[35] Ana-Andreea Stoica, Christopher Riederer, and Augustin Chaintreau. Algorithmic glass ceiling in social networks: The effects of social recommendations on network diversity, 2018.

[36] Jessica Su, Aneesh Sharma, and Sharad Goel. The effect of recommendations on network structure, 2016.

[37] Gergő Tóth, Johannes Wachs, Riccardo Di Clemente, Ákos Jakobi, Bence Ságvári, János Kertész, and Balázs Lengyel. Inequality is rising where social network segregation interacts with urban topology. *arXiv preprint arXiv:1909.11414*, 2019.

We present missing proofs and additional discussions in this appendix.

A Triadic Closure and Stochastic Block Models

Recall that for ease of presentation in the main text we state results for the case where $k = 2$. We remove this assumption and present generalized results below.

Lemma 6. *Given a network G resulting from an SBM with $k \geq 2$, the expected number of monochromatic wedges is:*

$$\sum_{i=1}^{k}\left(3\binom{n/k}{3}p^2(1-p) + (n-n/k)\binom{n/k}{2}q^2(1-p)\right) \tag{4}$$

while that of bichromatic wedges is

$$\sum_{i=1}^{k} 2(n-n/k)\binom{n/k}{2}pq(1-q). \tag{5}$$

Proof. The first term in Equation 4 results from the expected number of monochromatic wedges from a monochromatic triplets is where $\binom{n/k}{3}$ counts the number of monochromatic triplets of each type, 3 is the number of ways to choose the center of the wedge, and $p^2(1-p)$ counts the likelihood that such a triplet results in a wedge. The second term results from the expected number of monochromatic wedges from a bichromatic triplet, where $(n-n/k)\binom{n/k}{2}$ counts the number of bichromatic triplets and $q^2(1-p)$ corresponds to the likelihood that such a triplet results in a monochromatic wedge.

Likewise, for Equation 5, we note that $(n-n/k)\binom{n/k}{2}$ counts the number bichromatic triplets. Such a triplet results in a bichromatic wedge with probability $pq(1-q)$. □

We can similarly analyze the effect of triadic closure on network integration for $k \geq 2$.

Theorem 7. *Given an SBM with $k \geq 2$, there is a sufficiently large n such that triadic closure increases network integration if $p > q$.*

Proof. The fraction of bichromatic edges, $f(G)$, in this general case is:

$$\begin{aligned}
f(G) &= \frac{\frac{1}{2}\sum_i \frac{n}{k}\left(n-\frac{n}{k}\right)q}{\frac{1}{2}\sum_i \frac{n}{k}\left(n-\frac{n}{k}\right)q + \sum_i \binom{n/k}{2}p} \\
&= \frac{\frac{n}{2}\left(n-\frac{n}{k}\right)q}{\frac{n}{2}\left(n-\frac{n}{k}\right)q + \frac{k}{2}\left(\frac{n}{k}\right)^2 p} \\
&= \frac{\left(\frac{n^2}{2}-\frac{n^2}{2k}\right)q}{\left(\frac{n^2}{2}-\frac{n^2}{2k}\right)q + \frac{n^2}{2k}p} + o(1) \\
&= \frac{q(k-1)}{q(k-1)+p} + o(1)
\end{aligned}$$

Using the above lemma, we can simplify the fraction of bichromatic wedges to be

$$w(G) = \frac{(k-1)pq(1-q)}{(k-1)pq(1-q) + \frac{p^2(1-p)}{2} + \frac{k-1}{2}q^2(1-p)} + o(1)$$

As above, we are interested in the inequality $f(G) \leq w(G)$ for large enough n.

That is,

$$\frac{q(k-1)}{q(k-1)+p} \le \frac{(k-1)pq(1-q)}{(k-1)pq(1-q)+\frac{p^2(1-p)}{2}+\frac{k-1}{2}q^2(1-p)}$$

$$\frac{(k-1)^3q^3(1-p)}{2}+\frac{(k-1)qp^2(1-p)}{2} \le (k-1)qp^2(1-q)$$

$$(k-1)^2q^2(1-p)+p^2(1-p) \le 2p^2(1-q)$$

$$(k-1)^2q^2(1-p) \le p^2(1-2q+p)$$

We therefore look at the expression

$$p^2(1-2q+p)-(k-1)^2q^2(1-p).$$

For $k > 2$ and $q \in [0,1]$, this expression as roots at

$$q = \frac{p^2-\sqrt{-kp^4+2p^4+kp^2-p^2}}{(p-1)(k-1)}.$$

We therefore have that triadic closure preserves network integration for $q = \frac{p^2-\sqrt{-kp^4+2p^4+kp^2-p^2}}{kp-k-p+1}$ and increases integration for q less than this value.

The expression

$$p^2(1-2q+p)-(k-1)^2q^2(1-p)$$

simplifies to $p^2(1-p)k(2-k)$ for $q = p$. This value is positive if $k > 2$. Therefore, the insight that triadic closure increases network integration in the setting where nodes exhibit type-bias holds for $k \ge 2$.

□

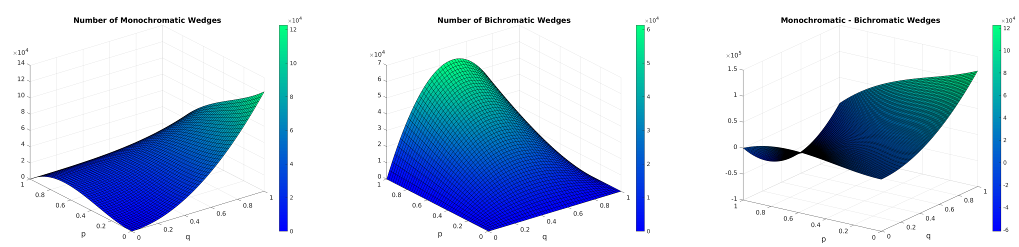

Figure 1: Wedge counts as we vary p and q for (a) monochromatic wedges, (b) bichromatic wedges, and (c) difference between monochromatic and bichromatic wedges for $n = 100$.

Note that for the case where $k = 2$, we find that $f(G) < w(G)$ if $p > q$ and $f(G) = w(G)$ if $p = q$; triadic closure increases network integration if nodes display homophily and it preserves network integration when there is no homophily. We note the following necessary subtlety in the argument for $k = 2$ which also extends to the general case. An SBM with $p > q$ always has $f(G) < 1/2$ for sufficiently large n, since $f(G) = q/(p+q) + o(1)$. Thus, one might imagine a strategy for proving this theorem that sought to show $w(G) \ge 1/2$, which would be sufficient. However, this is not always the case as we can see from Figure 1 – in particular, for low values of q and high values of p, there can be more monochromatic wedges than bichromatic wedges, and hence $w(G) < 1/2$. Given this observation, we must therefore focus on a more careful analysis of the relative sizes of $f(G)$ and $w(G)$.

B A Dynamic Network Formation Model

We present a heuristic argument to support Identity 1 in this section. First, we compare $f_R(t)$, $f_B(t)$, and $f(t)$. By running simulations on various parameters, we find that $f_R(t)$ and $f_B(t)$ both also converge to f^* for large enough t. This observation agrees with the intuition that the model proceeds

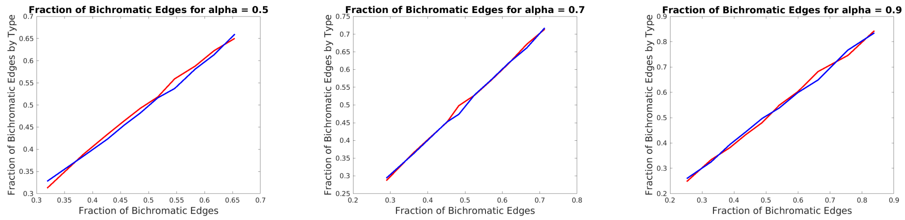

Figure 2: Values for $f_R(t), f_B(t)$, and $f(t)$ as N_S goes from 1 to 10 for $N_D = 10 - N_S$ and $N_F = 20$.

by adding a node of a random type then proceeding symmetrically regardless of the node type. Therefore, as the network approaches the equilibrium state, we can also expect the $f_R(t)$ and $f_B(t)$ to approach this value by symmetry.

We run the plots for Figure 3 above on $n = 500$ nodes and average over 10 trials. We note that these values on the fraction of bichromatic edges, including those disaggregated by type, converge to the expected values relatively quickly. And, most notably, $f_R(t)$ and $f_B(t)$ show comparable values for a range of α values. Note that we run these plots for a combination of $N_S + N_D$ and N_F values and the qualitative insights remain the same.

Lemma 8. *The fraction of bichromatic edges of $G(t)$ as t goes to infinity converges to:*

$$f(t) = \frac{N_D + (1-\alpha)N_F}{N_D + N_S + 2(1-\alpha)N_F}.$$

Proof. In equilibrium, $f(t+1) = f(t)$. Therefore:

$$\begin{aligned}
\frac{b(t)}{m(t)} &= \frac{b(t) + N_D + \alpha N_F f(t) + (1-\alpha)N_F(1-f(t))}{m(t) + N} \\
Nb(t) &= m(t)\left(N_D + \alpha N_F f(t) + (1-\alpha)N_F(1-f(t))\right) \\
Nf(t) &= N_D + \alpha N_F f(t) + (1-\alpha)N_F - (1-\alpha)N_F f(t) \\
Nf(t) &= N_D + (2\alpha - 1)N_F f(t) + (1-\alpha)N_F \\
f(t) &= \frac{N_D + (1-\alpha)N_F}{N + (1-2\alpha)N_F}
\end{aligned}$$

This last equality holds since $N = N_S + N_D + N_F$.

□

Rate of Convergence.

Finally, we find that we can also explicitly state how each of these parameters impact the rate at which the network converges to the equilibrium state for our dynamic network formation model.

At equilibrium, $f(t) = f(t+1)$. We are therefore interested in

$$\begin{aligned}
&f(t+1) - f(t) \\
&\quad = \frac{N_D + \alpha N_F f(t) + (1-\alpha)n_F(1-f(t))}{N_S + N_D + N_F} - f(t)
\end{aligned}$$

We solve the differential equation,

$$\frac{df(t+1)}{dt} = \frac{N_D + 2\alpha N_F f(t) + n_F(1-f(t))}{N_S + N_D + N_F} - f(t)$$

to get that,

$$f(t) = \frac{N_D + N_F - c \cdot \exp(-(t(N_F - 2\alpha N_F))/(N)}{N_S + N_D + 2(1-\alpha)N_F}$$

where c is some constant. In addition to impacting the equilibrium state, triadic closure also affects the rate of convergence to this state. This holds even for the case where $\alpha = 1$ when triadic closure preserves network integration.