CS 224W FINAL PROJECT: QUASIRANDOMNESS AND SIDORENKO'S CONJECTURE IN DIRECTED NETWORKS

NITYA MANI

1. Introduction

1.1. **Context.** A challenging and important question in network analysis is counting the number of copies of some *motif* B in some larger graph G as a way to featurize a graph for downstream learning tasks on the network, understand fundamental graph theoretic properties of G or make predictions about how far from random G is.

The associated fundamental extremal graph theory question is estimating the minimum number of copies of subgraph B in any graph G on a fixed number of vertices and edges. As a special case, one very famous class of graph theory questions is the set of Turán-type problems, asking how many edges a graph G on a fixed number of vertices must have to guarantee it has at least one copy of fixed subgraph B. Special cases (B as a triangle, clique) were resolved by Turán, Mantel, and dramatically generalized to an asymptotic answer for all non-bipartite fixed subgraphs B by the Erdős-Stone-Simonivits Theorem. Although a variety of upper bounds have been shown for bipartite B, a tight bound for all bipartite B has eluded mathematicians for over a century.

Questions about motif counts in networks can be very naturally posed as questions about the density of fixed subgraphs in larger graphs. We often understand graphs G on $E(G)$ vertices and $V(G)$ edges via their *edge density*, $p = E(G)/\binom{V(G)}{2}$. Here we consider more generally the B*-density* of a graph G, the fraction of injective vertex maps $\rho : V(B) \to V(G)$ that send edges to edges. Each such map ρ gives a distinct labeled copy of B in G. For fixed N and edge density p, we often wish to compute minimum possible B-density in graphs G. We obtain an upper bound on this quantity by taking $G = \mathcal{G}(N, p)$ to be an Erdos-Renyi random graph. In such G, the minimum possible B density is at most $p^{|E(B)|}$.

1.2. **Motivation.** With respect to this formulation, one of the primary motivating questions for all of extremal graph theory for the past several decades has been *Sidorenko's conjecture*, the surprising result that the above upper bound on the minimal B-density in graphs G is sharp when B is bipartite. More precisely, this conjecture jointly posed by Erdős-Simonivits [12] and Sidorenko [11] proposes that for any bipartite graph B on m edges, there exists some constant $\epsilon(B) > 0$ such that the number of copies of motif B in any graph G on N vertices (for sufficiently large N with edge density $p > N^{-\epsilon(B)}$) is at least $p^{|E(B)|}N^{|V(B)|}$, the expected number of copies of B in the Erdos-Renyi random graph $\mathcal{G}(N, p)$.

The above presentation of Sidorenko's conjecture immediately suggests its relevance to making sense of graphlet and motif counts in networks, and to understanding features of networks that may seem surprising at first glance. However, beyond the potential for Sidorenko's

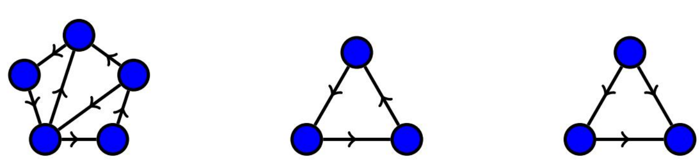

Figure 1. Digraph G on 5 vertices (left), $B = C_3^*$ (middle), tournament on 3 vertices B' (right). The B-density of G is $\frac{1}{10}$ and the B'-density is $\frac{1}{60}$.

conjecture to inform motif-based node featurization, it also has a wide variety of applications to random matrix theory, Markov chains, and understanding *quasirandomness*.

Quasirandom graphs were first studied by Thomason, and Chung-Graham-Wilson [2] who observed that a large number of properties that Erdos-Renyi random graphs satisfy are actually equivalent. Such properties can be used to understand one of the primary motivating questions in network analysis: "How close to random is a given network G?" Understanding deterministic graph constructions that have such properties (i.e. *quasirandom graphs*) can be useful as a benchmark when wondering whether features of a network are idiosyncratic or expected based on its fundamental characterization.

The notion of quasirandomness also leads to a strengthening of Sidorenko's conjecture. A graphlet B is *forcing* if a family of graphs $\{G_n\}_{n=1}^{\infty}$ is quasirandom if and only if the number of copies of B in G_n is asymptotically the number achieved in the Erdos-Renyi graphs of density matching G_n. As an example, we can take $B = C_4$ (a cycle of length 4), a motif shown to be forcing. C_4 is forcing is the statement that a family of graphs $\{G_n\}_{n=1}^{\infty}$ with edge density p is quasirandom (i.e. behaves like $\mathcal{G}(n,p)$ for most mathematical and computational purposes) if and only if it's C_4 density is roughly p^4. The *forcing conjecture*, initially posed by Skokan and Thoma [13] states that graphlets B are forcing if and only if they are bipartite and contain a cycle (showing these conditions are necessary is straightforward). The forcing conjecture would yield a short certificate of a graph behaving "like random."

While an extensive effort has gone in over the past decades to resolving parts of Sidorenko's conjecture, the forcing conjecture, and related extremal claims about networks, the analogues of these problems for directed or oriented networks have gone largely unstudied. In fact, apart from developing an analogous characterization of quasirandomness in directed graphs, as in [3], little work has been done to investigate extremal questions concerning directed motif counts. This problem is substantially more challenging, with very limited understanding of what directed analogues of the above two conjectures are likely to be true.

1.3. **Contribution Overview.** Here, we present an original characterization of a directed Sidorenko conjecture and a directed forcing conjecture and show necessary conditions on directed motifs to satisfy these results. We relate these characterizations to the undirected analogues and show the limitations of reductions of directed problems to undirected graphs[1].

1.4. **Outline.** We proceed as follows through this article. We begin by stating our main results in Section 2. In Section 3, we review notational preliminaries used throughout the article. In Section 4, we recall previous literature on undirected Sidorenko, forcing results, motif counting, and quasirandomness in directed networks. Armed with this background, we give a broader characterization of quasirandom directions in directed networks in Section 5. These results about quasirandom orientations enable us to state results about directed forcing motifs. In Section 6, we present a natural directed forcing conjecture and give context and motivation for Theorem 2.4 and Theorem 2.4. We tackle directed Sidorenko in Section 7. We set up a symmetric and asymmetric directed Sidorenko conjecture and present several relationships between the directed and undirected conjectures, motivating our major results Theorem 2.3 and Theorem 2.5. Finally, we discuss some remaining open problems ripe for future investigations, give applications of our work, and conclude in Section 8.

[1]I was approved to do a purely theoretical graph theory project by Prof. Leskovec and thus do not have a Github repository. However, I do illustrate the results with a few toy graphs pictured throughout and discuss applications of these theorems in Section 8.

2. Statement of Results

We show several state-of-the art results, making substantial progress towards understanding directed motif counts in "like-random" directed graphs and lower bounds on the counts of families of directed motifs in all directed graphs. All of our novel contributions are collected below, but relevant notation may only be introduced later in the article (c.f. Section 3); we repeat the results in context after showing the necessary intermediate results and defining all relevant quantities precisely.

We first give an expanded characterization of quasirandom orientation in directed graphs of independent computational interest.

Theorem 2.1. *For a digraph $G = (V, E)$ on n vertices and $m = \Omega(n^2)$ edges with underlying undirected graph H, the following are equivalent:*

(1) $\tau(G) = o(m)$
(2) $\tau^*(G) = o(m)$
(3) $N(C_4^{\rightarrow}, G) = \left(\frac{1}{8} + o(1)\right) N(C_4, H)$, *where* $C_4^{\rightarrow} = \{(v_1, v_2), (v_3, v_2), (v_3, v_4), (v_1, v_4)\} \subset E$ *for vertices* $v_1, v_2, v_3, v_4 \in V$
(4) *For any labeling L of V and all B,* $N_L(B, G) = \left(2^{-|E(B)|} + o(1)\right) N_L(B, H)$
(5) *For any even* $k \geq 4$, $E_k(G) = \left(\frac{1}{2} + o(1)\right) N_L(C_k, H)$.
(6) *For any even* $k \geq 4$, $\mathrm{Tr}(A(G)^k) = o(\mathrm{Tr}(A(H)^k)))$
(7) $|\lambda_1(G)| = o(|\lambda_1(H)|)$

If any of these conditions is satisfied, G has quasirandom direction *with respect to H.*

We give a broad, infinite family of directed motifs B such that counting the copies of B in *any* directed graph G completely characterizes whether or not any tournament (orientation of a complete graph) has spectral and structural properties that cause it to behave "like random" for almost any computational purpose.

Theorem 2.2. *If $B = (V, E)$ is a transitive directed graph such that the underlying graph $\overline{B}$ satisfies the asymmetric forcing property then for any tournament G, G has quasirandom direction iff*

$$t_B(G) = (\mu(B) + o(1))$$

We also give a broad infinite family of directed motifs B that are overrepresented in *all* tournaments, relative to randomly orienting the edges of a complete graph:

Theorem 2.3. *Let $B = (B_1 \sqcup B_2, F)$ be any bipartite digraph such that for all $e = (b_1, b_2) \in F$, $b_1 \in B_1, b_2 \in B_2$ with underlying undirected graph $\overline{B}$. Then for any tournament $G = (V, E)$, if $\overline{B}$ satisfies asymmetric Sidorenko's conjecture, we have a Sidorenko-style bound:*

$$t_B(G) \geq \mu(B)$$

More generally, we are also able to give strong necessary conditions for any directed motif to have a directed Sidorenko property or be forcing for general directed graphs:

Theorem 2.4. *If a digraph $B = (V, E)$ that satisfies $|V| = b > (4(1 + \epsilon))^{1+1/\epsilon}$ and $|E| \geq (1 + \epsilon)b$ for any fixed $\epsilon > 0$ is not transitive, it is not forcing.*

We show a necessary condition for a directed motif to be overrepresented:

Theorem 2.5. *Any digraph B satisfying the directed Sidorenko property must be transitive.*

3. Notational Preliminaries

Throughout, all directed graphs (abbreviated as *digraphs*) are assumed to be unweighted, oriented graphs (i.e. with no parallel or antiparallel edges and no self-loops). Unless, otherwise specified, we let $G = (V, E)$ be a digraph with $|V| = n$ vertices and $|E| = m$ edges. We will be interested in the undirected graph associated to G:

Definition 3.1. For a digraph $G = (V, E)$, we define the *underlying undirected graph* $H = (V, F)$ where for each edge $v \to w = (v, w) \in E$ we have an undirected edge $(v, w) \in F$.

For a digraph $G = (V, E)$ and vertices $a, b \in V$, let $(a, b) \in E$ be the edge directed $a \to b$. For subsets $A, B \subset V$, let

$$e(A, B) = |\{e = (a, b) \in E \mid a \in A,\, b \in B\}|$$

denote the number edges directed from vertices in A to vertices in B and vice versa for $e(B, A)$. For a vertex $v \in V$, let $d^+(v)$ be the *indegree* of a vertex $v \in V$, i.e.

$$d^+(v) = |\{w \in V \mid (w, v) \in E\}|,$$

and similarly let $d^-(v)$ denote the *outdegree* of v. We let $d^+(v, S)$, be the *indegree into* S of v:

$$d^+(v, S) = |\{w \in S | (w, v) \in E\}|$$

and similarly let $d^-(v, S)$ be the *outdegree into* S of v.

Definition 3.2. Given a graph $G = (V, E)$, for two subsets $A, B \subset V$ we define

$$\tau(G) := \max_{A,B \subset V} (e(A, B) - e(B, A)),$$

$$\tau^*(G) := \max_{A,B \subset V, A \cap B = \emptyset} (e(A, B) - e(B, A)).$$

We can also similarly define the *maximal edge difference for partitions*:

$$\tau_\sqcup(G) = \max_{A \cup B = V, A \cap B = \emptyset} (e(A, B) - e(B, A))$$

Note τ^* is not always achieved by a partition although $\tau_\sqcup(G) \leq \tau^*(G)$.

A motivating question is counting directed motifs in graphs, which we do in 2 ways:

Definition 3.3. For digraph $G = (V, E)$ with underlying undirected graph H and a digraph B (with undirected underlier $\overline{B}$), let $N(B, H)$, be the number of copies of $\overline{B}$ in H and $N(B, G)$ be number of copies of B in G. For a labeling $L : V \to [n]$ of the vertices, let $N_L(B, H)$ be the number of labeled copies of $\overline{B}$ in H and let $N_L(B, G)$ be the number of labeled copies of B as a subgraph of G. In other words, N counts graphlets once and N_L counts graphlets as many times as there are automorphisms.

In the introduction, we articulated Sidorenko's conjecture and the forcing conjecture in the language of motif counts and densities. In practice, we will work with a far more flexible and useful characterization of these conjectures in the language of homomorphism densities (we can do this at the same time for directed and undirected networks):

Definition 3.4. A *graph homomorphism* $B \to G$ is a map $\rho : V(B) \to V(G)$ such that if $(v, w) \in E(B)$, then $(\rho(v), \rho(w)) \in E(G)$ (i.e. ρ maps edges to edges). The *homomorphism density* of B in G, denoted $t_B(G)$ is the fraction of vertex maps that are homomorphisms:

$$t_B(G) = \frac{h_B(G)}{|V(G)|^{|V(B)|}},$$

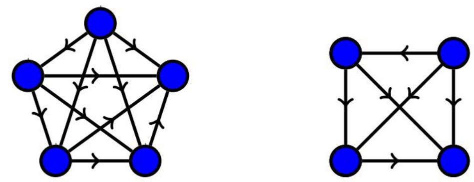

Figure 2. Transitive 5-vertex (left) and 4-vertex tournament (right)

where $h_B(G)$ is the number of homomorphisms $B \to G$.

Homomorphisms are not necessarily injective, multiple vertices of B can map to the same target vertex in G. The classical Sidorenko's conjecture is just the following:

Conjecture 3.5 (Sidorenko's Conjecture)**.** *For every bipartite undirected graph A with m edges and every undirected graph H (we denote an edge by K_2),*

$$t_A(H) \geq t_{K_2}(H)^{|E(A)|}$$

We consider oriented graphs in general, but many of our results focus on a specific family of directed graphs, the so-called *tournaments* (*directed cliques*):

Definition 3.6. A *tournament* on n vertices is a digraph $T = (V, E)$ with $|V| = n$ such that for every pair of vertices $\{v, w\}$ exactly one of $(v, w), (w, v) \in E$.

4. Previous Work

4.1. **Sidorenko's conjecture.** Beyond the initial observations that Sidorenko's conjecture held for cycles, one of the most substantial steps was taken by Conlon, Fox, and Sudakov in "An approximate version of Sidorenko's conjecture" [4]. First, they showed Conjecture 3.5 held when $A = (V_1 \sqcup V_2, E)$ was a bipartite undirected graph with a vertex with an edge to every node in the other part. They also proved an approximate version of Conjecture 3.5 and the forcing conjecture for a family of bipartite graphs A.

4.2. **Quasirandom tournaments.** The above literature deals exclusively with extremal results for motif counts and quasirandomness properties for *undirected graphs.* The study of such questions in directed graphs is severely limited. Chung and Graham, who first introduced quasirandomness in undirected graphs, also gave a characterization of quasirandom tournaments (directed cliques). They posed a natural question: "Given a tournament, how is it possible to tell if the tournament and its properties behave like-random or have some special features?" They gave an analysis of several equivalent properties that are all shared by random tournaments in [3] (i.e. given by uniformly at random picking one of the two possible orientations of each edge in a clique).

They gave 11 equivalent properties that provide a short certificate that a given explicit tournament has "random-like" behavior in contrast to checking such properties for an instantiation of a random tournament. This work was later generalized in [7] to give spectral characterizations of quasirandom tournaments. In both cases, progress was limited to quasirandomness for this special family of directed graphs without describing forcing or Sidorenko-style properties that may or may not hold for directed motifs in tournaments.

4.3. **Quasirandom orientation.** In 2013, Griffiths gave a presentation of quasirandom orientations on a general directed graph in [6]. He focused on showing analogues of the properties described in [3] for oriented and partially oriented graphs.

In the process, Griffiths was able to show the first forcing results for directed graph, by showing that two orientations of an undirected 4-cycle were forcing (one where two vertices have out-degree 2 and the other 2 have in-degree 2 ($C_4^{\rightarrow}$), and another orientation given as a directed cycle with a single flipped edge). In the process, he also concluded that the other two (up to isomorphism) orientations of a 4 cycle would not satisfy a forcing conjecture. This limited result highlights how much more difficult the problems of forcing and Sidorenko style bounds are for directed graphs than for undirected graphs. This paper is limited to showing forcing for these two very specific subgraphs and falls short of any Sidorenko-style analysis or forcing claims even for slightly larger cycles (such as for orientations of a 6-cycle).

5. Quasirandom Directions

Throughout, we let $G = (V, E)$ be a digraph on $|V| = n$ vertices and $|E| = m$ edges. We use $\tau(G)$ and labeled counts of copies of subgraphs to characterize graphs with quasi-random directions, as introduced in [3] and extended to all digraphs in [6].

We consider properties that a dense digraph $G = (V, E)$ on n vertices and $m = \Omega(n^2)$ vertices might satisfy. We use the asymptotic $o(\cdot)$ notation loosely. If $P = P(o(1)), Q = Q(o(1))$, then $P \implies Q$ means that for each $\epsilon > 0$, for some sufficiently large $n > N(\epsilon)$, there exists δ so that if G satisfies $Q(\delta)$ then it satisfies $P(\epsilon)$.

Definition 5.1. For a digraph $G = (V, E)$ on n vertices, the adjacency matrix A is an $n \times n$ adjacency matrix with rows and columns indexed by vertices so that

$$A_{uv} = \begin{cases} 1 & (u, v) \in E \\ -1 & (v, u) \in E \\ 0 & \text{else} \end{cases}$$

These definitions will allow us to give an expanded characterization of Theorem 2.1, replicated below. We defer the proof of Theorem 2.1 to Appendix A

Theorem. *For a digraph $G = (V, E)$ on n vertices and $m = \Omega(n^2)$ edges with underlying undirected graph H, the following are equivalent:*

(1) $\tau(G) = o(m)$
(2) $\tau^*(G) = o(m)$
(3) $N(C_4^*, G) = \left(\frac{1}{8} + o(1)\right) N(C_4, H)$, *where* $C_4^* = \{(v_1, v_2), (v_3, v_2), (v_3, v_4), (v_1, v_4)\} \subset E$ *for vertices* $v_1, v_2, v_3, v_4 \in V$.
(4) *For any labeling L of V,* $N_L(B, G) = \left(2^{-|E(B)|} + o(1)\right) N_L(B, H)$
(5) *For any even* $k \geq 4$, $\mathrm{Tr}(A(G)^k) = o(\mathrm{Tr}(A(H)^k)))$
(6) $|\lambda_1(G)| = o(|\lambda_1(H)|)$

If any of these conditions is satisfied, G has quasirandom direction *with respect to H.*

Being quasirandom endows a digraph with a tremendous amount of structure. As an illustration, in Appendix A we also show the implications of quasirandomness for a directed graph being *almost-balanced*: $\sum_{v \in V} |d^+(v) - d^-(v)| = o(m)$.

6. Forcing Oriented Graphs

We recall the classical definition of *quasirandom undirected graphs* via *forcing subgraphs*, related to the characterization of Theorem 2.1

Definition 6.1. A sequence $(H_n : n = 1, 2, \ldots)$ of undirected graphs is called *quasirandom with density* p (where $0 < p < 1$) if, for every graph A,

$$t_A(H_n) = (1 + o(1))p^{|E(A)|},$$

$$t_A(H) = \frac{h_A(H)}{|V(H)|^{|V(A)|}}$$

is the fraction of mappings $f : V(A) \to V(H)$ which are homomorphisms.

The above definition gives rise to p-forcing subgraphs, individual subgraphs that guarantee quasirandomness. This is made more precise below:

Definition 6.2. A graph A is *p-forcing* if a sequence of undirected graphs H_n is quasirandom with density p only if

$$t_A(H_n) = (1 + o(1))p^{|E(A)|}$$

A graph A is said to be *forcing* if it is p-forcing for all p.

Conjecture 6.3 (Forcing Conjecture)**.** *An undirected graph A is forcing if and only if it is bipartite and contains a cycle.*

We also consider the directed analogue of *forcing subgraphs*, as in [4]. To do this, we will first need to understand the symmetries of digraphs:

Definition 6.4. On a vertex labeled digraph B with underlying undirected graph H, we define $\mu(B)$ as the fraction of directed graphs with underlying graph H isomorphic to B (i.e. the fraction of orientations of H that yield digraphs C that there exists a vertex isomorphism $V(C) \cong V(B)$ mapping edges to edges). Note that

$$\mu(B) = \frac{\text{Aut}(H)}{\sigma(B) \cdot 2^{|E(B)|}}.$$

where $\text{Aut}(H)$ counts the labeled automorphisms of undirected graph H and $\sigma(B)$ counts the number of symmetries of digraph B.

Example. C_6 oriented so all edges go from one part to the other, termed $C_6^{\to}$, has $\mu(C_6^{\to}) = 2/2^6 = 1/32$, whereas C_6 oriented as two length three paths has $\mu(C_6^p) = 6/2^6 = 3/32$.

Definition 6.5. For any digraph $G = (V, E)$, a digraph B is *forcing* if G is quasirandom only if

$$t_B(G) = (\mu(B) + o(1))\, t_{K_2}(G)^{|E(B)|}$$

We say that a digraph is *forcing for a family of digraphs* $(H_n)_{n=1}^{\infty}$ if for any digraph $G_n = (V, E)$ with underlying undirected graph H_n as $n \to \infty$, G is quasirandom only if

$$t_B(G_n) = (\mu(B) + o(1))\, t_{K_2}(H_n)^{|E(B)|}$$

This setup allows us to show Theorem 2.4 to give a necessary condition for a directed motif to be forcing, replicated below. We defer the proof to Appendix B.

Theorem. *If a digraph B that satisfies $|V(B)| = b > (4(1+\epsilon))^{1+1/\epsilon}$ and $|E(B)| \geq (1+\epsilon)b$ for any fixed $\epsilon > 0$ is not transitive, it is not forcing.*

Further, if the underlying undirected graph $\overline{B}$ of transitive digraph B satisfies a stronger *asymmetric forcing property*, then for any tournament G, G has quasirandom direction iff

$$t_B(G) = (\mu(B) + o(1))$$

We prove this result (Theorem 2.2 and discuss asymmetric forcing further in Appendix B.

7. The Directed Sidorenko Conjecture

Our characterization of quasirandom graphs is closely tied to the famous Sidorenko conjecture, stated via graph homomorphisms in Conjecture 3.5. We consider the analagous questions for digraphs. Given undirected graph $H = (V, F)$, a random *orientation* on H, $G = (V, E)$ is given by taking for each $(v, w) \in F$ exactly one of $(v, w), (w, v)$ to be in E uniformly at random. Conversely, for any digraph $G = (V, E)$, let $H = (V, F)$ denote the underlying undirected graph.

Definition 7.1 (Directed Sidorenko)**.** We define two Sidorenko-style properties for digraphs B based on one of (7.1) and (7.2) for every digraph $G = (V, E)$:

$$t_B(G) \geq \mu(B) t_{K_2}(G)^{|E(B)|} + o(1) \tag{7.1}$$

$$t_B(G) \geq \mu(B) t_{\overline{B}}(H) + o(1) \tag{7.2}$$

A digraph B has the *directed Sidorenko property* if for all digraphs $G = (V, E)$, (7.2) holds.

Note that (7.2) implies (7.1) if the underlying graph satisfies Sidorenko's conjecture. Thus, the above definition captures a natural directed analogue of the Sidorenko property:

Proposition 7.2. *If an undirected graph A does not have the Sidorenko property, then for all orientations B of A, B does not satisfy* (7.1)*.*

Proof. Suppose undirected A does not have the Sidorenko property. Then, for some family of graphs (H_n) for $H = H_n$ (n sufficiently large), we have

$$t_A(H) < t_{K_2}(H)^{|E(A)|}$$

Suppose we randomly orient the edges of H to obtain G. Let B be a fixed orientation of A. The expected number of copies of B is $\mu(B) t_A(H)$, implying that

$$\mathbb{E}[t_B(G)] = \mu(B) t_A(H) < \mu(B) t_{K_2}(H)^{|E(A)|} = \mu(B) t_{K_2}(G)^{|E(B)|}.$$

Therefore, there exists some digraph G such that $t_B(G) < \mu(B) t_{K_2}(G)^{|E(B)|}$, so B does not satisfy (7.1). Since this holds for all orientation of A, we obtain the desired result. ■

As in Section 6, we can extend our setup to an asymmetric directed Sidorenko property. We recall the classical undirected characterization and define a directed bound:

Definition 7.3. Bipartite undirected graphs $\overline{B} = (V_1 \sqcup V_2, E)$, $|E| = m$ and $H = (U_1 \sqcup U_2, F)$ with edge density $p = \frac{|F|}{|U_1||U_2|}$ satisfy the *asymmetric Sidorenko property* if the density of homomorphisms $f : V(\overline{B}) \to V(G)$ such that $f(V_i) \subset U_i$ for $i = 1, 2$ is at least p^m, i.e.

$$t_{\overline{B}}(H) \geq p^m.$$

Bipartite directed graphs $B = (V_1 \sqcup V_2, E)$, $|E| = m$ and $G = (U_1 \sqcup U_2, F)$ with edge density $p = \frac{|F|}{|U_1||U_2|}$ satisfy the *directed asymmetric Sidorenko property* if the density of maps $f : V(B) \to V(G)$ such that $f(V_i) \subset U_i$ for $i = 1, 2$ which are homomorphisms is at least $p^m \mu(B)$, in other words, $t_B(G) \geq p^m \mu(B)$.

This characterization suggests a possible reduction that we give, showing that orientations of undirected graphs that satisfy the asymmetric Sidorenko's conjecture satisfy the Sidorenko style-bound in tournaments of Theorem 2.3, replicated below and proved in Appendix C:

Theorem. *Let $B = (B_1 \sqcup B_2, F)$ be any bipartite digraph such that for all $e = (b_1, b_2) \in F$, $b_1 \in B_1, b_2 \in B_2$ with underlying undirected graph $\overline{B}$. Then for any tournament $G = (V, E)$, if $\overline{B}$ satisfies asymmetric Sidorenko's conjecture, we have a Sidorenko-style bound:*

$$t_B(G) \geq \mu(B).$$

We also observe in Theorem 2.5 (proved in Appendix C) that any digraph B with the directed Sidorenko property must be transitive.

8. Discussion

8.1. Applications.

Progress towards the directed Sidorenko and forcing conjecture such as the results of Section 2 are of substantial interest to the computer science community. We highlight three of numerous potential applications of such results.

8.1.1. *Graphlet featurization.* Making predictions in directed graphs often requires pre-processing for featurization. Thus, learning effective whole-graph embeddings is of considerable interest to those wishing to perform downstream machine learning tasks on graphs.

One common way of learning such embeddings is by embedding a graph into a feature space with each entry comprising a motif count for a set of small graphlets, as discussed in lecture. Often, we wish to understand the significance score of these graphlet counts in relation to random graphs; we understand properties of the graph based on how positive or negative such significance scores are. The directed Sidorenko conjecture would have the surprising consequence that whichever directed graphlets satisfy a directed Sidorenko property will never have negative expected significance score when comparing to an Erdős-Renyi random graph. This implies that featurizations with such motifs lose more information than can be expected, and inferring properties without this knowledge can lead to increased weight being put on motif counts that are not as surprising as the Z-score may naively suggest.

8.1.2. *Deterministic randomness.* In many practical engineering problems, including circuit design, building telecommunication networks, and understanding biomedical networks, we often wish to have a reference "like-random" directed graph that we can guarantee has desired good properties. Understanding pseudorandom directed graphs as we do in Theorem 2.1 provides short certificates for a deterministic graph to have "like random" behavior. In addition, this analysis yields a deterministic method to construct "like-random" graphs quickly with properties that make them good null models for experiments on real-world networks. Conversely, one of the necessary and sufficient conditions of quasirandom is given by counts of forcing directed motifs. Knowing that a directed network is quasirandom thus gives a rapid way to approximately enumerate directed subgraphs.

8.1.3. *Motifs in tournaments.* In addition to partial progress towards the general directed Sidorenko and forcing conjectures, our progress above shows Sidorenko-style bounds for tournaments. Tournaments occur all over the Internet and world, from athletics to auctions to Internet competitions, and our results enable such complete networks to be far better understood than they historically were. Specific applications include judging if participants in a

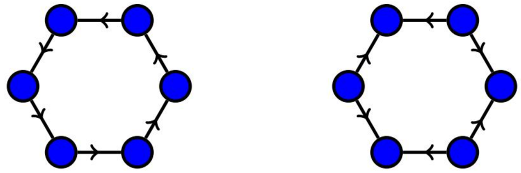

Figure 3. C_6^* (left) and $C_6^{\rightarrow}$ (right).

chess or other tournament are over-qualified or cheating by assessing tournament randomness and identifying sport-specific idiosyncrasies in round-robins.

Beyond these instances, understanding motif counts in directed graphs and short certificates of directed networks having "like-random" properties is of substantial interest to computational biologists, network scientists, and a host of other computer scientists.

8.2. **Open Questions.** The directed Sidorenko and forcing conjectures are far more poorly understood than their undirected analogues. Consequently, the progress highlighted in Section 2 represents state-of-the-art research results, implying that a number of relatively simple-seeming fundamental questions still remain open in the field. We introduce one such example:

Definition 8.1. Let *directed cycle* C_r^* be the directed graph on r vertices $v_1, \dots v_r$ comprising edges $(v_1, v_2), \dots (v_{r-1}, v_r), (v_r, v_1)$. Let $C_{2r}^{\rightarrow} = (V \sqcup W, E)$ be the bipartite digraph with underlying undirected graph C_{2r} so that for $e = (v, w) \in E$, $v \in V$ and $w \in W$.

One major result in the area of motif counting is that all even-length undirected cycles satisfy the undirected Sidorenko conjecture. However the analagous directed question is still wide open, even for the very specific case of computing which (if any) orientations of a 6-cycle are systematically overrepresented:

Question 8.2. Does the directed motif $C_6^{\rightarrow}$, consisting of an orientation of a 6-cycle where consecutive edges go in opposite directions, satisfy the directed Sidorenko conjecture? Does any orientation of a 6-cycle satisfy the directed Sidorenko conjecture?

More broadly, the broad problem our above results make progress towards is the following:

Question 8.3. Which directed motifs satisfy a directed Sidorenko and/or forcing conjecture?

Several of the methods highlighted in our partial results seem promising for making forward progress towards this motivating question. Further, they highlight potential reductions from directed graph questions to associated questions about undirected graphs where we can leverage a better understanding of graphlet counts.

Acknowledgements

I would like to thank my thesis advisor Professor Jacob Fox for his invaluable help in learning probabilistic graph theory. I would also like to thank him for several suggestions and references along the research process. I would also like to thank Zoe Himwich, with whom I have been collaborating on two other graph theory projects and who has helped me understand a variety of techniques in extremal graph theory. I would also like to thank Professor Leskovec and the CS 224W teaching team for their support.

References

[1] O. Amini, S. Griffiths, and F. Huc, Subgraphs of weakly quasi-random oriented graphs, *SIAM J. Discrete Math.* **25** (2011), 234–259.
[2] F. Chung, R. Graham, R. Wilson, Quasi-random graphs *Combinatorica* (1989), 345–362.
[3] F. Chung, R. Graham, Quasi-random tournaments *J. Graph Theory* (1991), 173–198.
[4] D. Conlon, J. Fox, and B. Sudakov, An approx. version of Sidorenko's conjecture, *Geom. Fun. Anal.* **20** (2010).
[5] D. Dellamonica et al., Tree-minimal graphs are almost regular, *J. Comb.*, **3** (2012), 49–62.
[6] S. Griffiths, Quasi-random oriented graphs, *J. Graph Theory* **74** (2013), 198–209.
[7] S. Kalyanasundaram and A. Shapira, A note on even cycles and quasirandom tournaments, *J. Graph Theory* (2013), 260–266.
[8] J. Fox, P. Keevash, and B. Sudakov, Directed graphs w/o short cycles, *Comb. Prob. Comp.* **19** (2010), 285–301.
[9] D. Mubayi and J. Verstraete, Counting trees in graphs, *E. J. Comb.* **23** (2016), 3–39.
[10] J. Shearer, A note on bipartite subgraphs of triangle-free graphs. *Rand. Struct. Alg.* **3** (1992), 223–226.
[11] A. Sidorenko, A correlation inequality for bipartite graphs. *Graph. Comb.* **9** (1993), 201–204.
[12] M. Simonovits, Extremal graph problems, degenerate extremal problems and super-saturated graphs. *Prog. Graph Theory* (1984), 419–437.
[13] J. Skokan, L. Thoma, Bipartite subgraphs and quasi-randomness. *Graphs Combin.* (2004), 255–262.

Appendix

Appendix A. Quasirandom and Almost-Balanced Directed Graphs

We will show the equivalences of Theorem 2.1 in a series of lemmas that follow, leveraging the results of [6, 7] that show some of the equivalence directions.

Lemma A.1. *For any graph $G = (V, E)$, $\tau^*(G) \le \tau(G) \le 3\tau^*(G)$.*

Proof. By construction $\tau(G) \ge \tau^*(G)$. Let

$$A^+, B^+ = \text{argmax}_{A,B \subset V}(e(A, B) - e(B, A))$$

so that

$$\tau(G) = e(A^+, B^+) - e(B^+, A^+) \ge 0$$

Let $J = A^+ \cap B^+$. If $J = \emptyset$, then $\tau(G) = \tau^*(G) \le 3\tau^*(G)$ and we are done. Else, we have that

$$\begin{aligned}
\tau(G) &= e(A^+, B^+) - e(B^+, A^+) \\
&= (e(A^+\backslash J, B^+\backslash J) + e(J, B^+\backslash J) + e(A^+\backslash J, J) + e(J, J)) \\
&\quad - (e(B^+\backslash J, A^+\backslash J) + e(J, A^+\backslash J) + e(B^+\backslash J, J) + e(J, J)) \\
&= (e(A^+\backslash J, B^+\backslash J) + e(J, B^+\backslash J) + e(A^+\backslash J, J)) \\
&\quad - (e(B^+\backslash J, A^+\backslash J) + e(J, A^+\backslash J) + e(B^+\backslash J, J)) \\
&= (e(A^+\backslash J, B^+\backslash J) - e(B^+\backslash J, A^+\backslash J)) + (e(J, B^+\backslash J) - e(B^+\backslash J, J)) \\
&\quad + (e(A^+\backslash J, J)) - (e(J, A^+\backslash J)) \\
&\le 3\tau^*(G)
\end{aligned}$$

■

To obtain a spectral characterization of quasirandomness, we will count even-switch k-cycles, defined below:

Definition A.2. For a digraph $G = (V, E)$, call an k-tuple of vertices $(v_1, \dots, v_k)$ an *even-switch k-cycle* if for $i = 1, \dots k$ (letting $v_{k+1} := v_1$) exactly one of $(v_i, v_{i+1}), (v_{i+1}, v_i) \in E$ and further $(v_{i+1}, v_i) \in E$ for an even number of i. Let $E_k(G)$, be the number of distinct even-switch k-cycles in G with respect to a labeling of V. Analogously, we can define an *odd-switch k-cycle*, and let $O_k(G)$ be the number of labelled odd-switch k-cycles in G.

In the above definition, we will assume that if $(v_1, \dots, v_k)$ is an even-switch k-cycle, then $(v_i, \dots v_{i+k \pmod k})$ defines the same even-switch k-cycle.

Lemma A.3. *For a digraph $G = (V, E)$ with underlying undirected graph H and some labelling L of V, for any even integer $k \ge 4$,*

$$\text{Tr}(A^k) = 2E_k(G) - N_L(C_k, H)$$

Thus,

$$E_K(G) = \left(\frac{1}{2} + o(1)\right) N_L(C_k, H) \iff \text{Tr}(A^k) = o(\text{Tr}(A(H)^k))$$

Proof. We follow the argument of [7]. Note that the (v, v) entry of A^k is the number of even-switch k-cycles with vertex v minus the number of odd-switch k-cycles (defined analogously) with vertex v. Thus, $\mathrm{Tr}(A^k) = E_k(G) - O_k(G)$. Note that

$$E_k(G) + O_k(G) = N_L(C_k, H) = \mathrm{Tr}(A(H)^k)$$

where $A(H)$ is the adjacency graph of H, defined as usual. This gives the desired equality:

$$\mathrm{Tr}(A^k) = 2E_k(G) - N_L(C_k, H) = 2E_k(G) - \mathrm{Tr}(A(H)^k)$$

In particular, this implies that

$$E_k(G) = \frac{1}{2}\left(\mathrm{Tr}(A^k) + N_L(C_k, H)\right)$$

so $E_K(G) = (\frac{1}{2} + o(1))N_L(C_k, H)$ if and only if $\mathrm{Tr}(A^k) = o(N_L(C_k, H)) = o(\mathrm{Tr}(A(H)^k))$. ■

Proof of Theorem 2.1. (1) $\iff$ (2) follows immediately from Lemma A. (5) $\implies$ (4) by setting $B = C_4$ and considering all possible valid labellings of a directed C_4 as a subgraph of G. Theorem 1.1 of [6] shows that (1) $\implies$ (5) and (4) $\implies$ (1), so (1) $\iff$ (4) $\iff$ (5). Lemma A.3 shows that (6) $\iff$ (7). (6) $\iff$ (8) and (6) $\iff$ (1) follow as noted in [7] applying the result of [6]. (8) $\iff$ (1) follows as in [6] noting that $m = \Theta(n^2)$ by assumption. ■

Quasirandom graphs have a number of beautiful properties including being *almost-balanced*:

Proposition A.4. *If digraph $G = (V, E)$ has quasirandom direction with respect to underlying undirected graph H, then G is almost balanced:*

$$\sum_{v \in V} |d^+(v) - d^-(v)| = o(m)$$

Proof. Note that if G has quasirandom direction with respect to H, then $\tau^*(G) = o(m)$. We follow the argument of [3]. Suppose to the contrary that G was not almost balanced. Then, we have some $W \subset V$ of $|W| = \epsilon n$ such that for each $v \in W$,

$$d^+(v) > \frac{d^+(v) + d^-(v)}{2} + \epsilon n \implies d^+(v) > d^-(v) + 2\epsilon n$$

Let B range over all subsets of $V\backslash W$ of size $n/2-|W|$. There are $\binom{n-|W|}{n/2-|W|}$ such choices of B. Thus, there exists some set A so that

$$\begin{aligned}
\tau^*(G) &\geq e(A\cup W,(A\cup W)^c)-e((A\cup W)^c,A\cup W)\\
&=\sum_{v\in A\cup W} d^+(v)-d^-(v)\\
&\geq \frac{1}{\binom{n-|W|}{n/2-|W|}}\sum_{B\subset V\backslash W||B|+|W|=n/2}\ \sum_{v\in B\sqcup W} d^+(v)-d^-(v)\\
&\geq \frac{1}{\binom{n-|W|}{n/2-|W|}}\binom{n-|W|-1}{n/2-|W|-1}e(W,W^c)-e(W,W^c)\\
&=\frac{n/2-|W|}{n-|W|}(e(W,V)-e(V,W))\\
&=\frac{1/2-\epsilon}{1-\epsilon}|W|2\epsilon n\\
&\geq \frac{2}{3}\epsilon^2 n^2
\end{aligned}$$

assuming that $\epsilon<\frac{1}{4}$, which gives a contradiction, since $\tau^*(G)=o(m)$, but $\frac{2}{3}\epsilon^2n^2=\Omega(m)$, since $m=O(n^2)$. Thus G is almost balanced. ∎

While a graph being almost balanced is not sufficient to guarantee quasi-randomness, it does show that edges are balanced around partitions of the graph:

Proposition A.5. *If a digraph G is almost balanced, then $\tau_\sqcup(G)=o(m)$.*

Proof. Suppose G is almost balanced. Then, for all $\epsilon>0$, for all but $\epsilon\frac{m}{n}$ vertices, we have that $|d^+(v)-d^-(v)|<\epsilon\frac{m}{n}$. Consider a partition $V=A\sqcup B$. We have that

$$\begin{aligned}
e(A,B)-e(B,A) &= e(A,A)+e(A,B)-e(A,A)-e(B,A)\\
&=e(A,V)-e(V,A)\\
&=\sum_{v\in A} d^+(v)-d^-(v)\\
&\leq \sum_{v\in A}|d^+(v)-d^-(v)|\\
&\leq \epsilon\frac{m}{n}\cdot n+|A|\cdot\epsilon\frac{m}{n}\\
&\leq 2\epsilon\frac{nm}{n}\\
&\leq 2\epsilon m
\end{aligned}$$

Thus $\tau_\sqcup(G)\leq 2\epsilon m$, and since this holds for all $\epsilon>0$, we obtain the desired result. ∎

Appendix B. Proof of Theorem 2.2 and Theorem 2.4

Here, we complete the proofs of our two major results making progress towards understanding directed motifs that have a forcing property.

We leverage a stronger version of the forcing property described in the article, as in [4].

Definition B.1. An bipartite undirected graph $\overline{B} = (V_1 \sqcup V_2, E)$ satisfies the *asymmetric forcing property* if for every bipartite $H = (U_1 \sqcup U_2, F)$, H is quasirandom if and only if

$$t_{\overline{B}}(H) = (1 + o(1))p^{|E(\overline{B})|},$$

where $p = \frac{|F|}{|U_1||U_2|}$ is the edge density of H. A transitive bipartite directed graph $B = (V_1 \sqcup V_2, E)$ satisfies the *directed asymmetric forcing property* if, for every $G = (U, F)$, G is quasirandom iff

$$t_B(G) = (\mu(B) + o(1))p^{|E(B)|},$$

where $p = \frac{|F|}{|U_1||U_2|}$ is the edge density of G.

Asymmetric forcing enables us in Theorem 2.2 (below) to give a broad family of directed motifs B, such that counting copies of B characterizes whether or not any tournament has spectral and structural properties that cause it to behave like random. We prove this in Appendix B

Proof of Theorem 2.2. First we fix a transitive directed graph $B = (V, E)$ with underlying undirected graph $\overline{B}$. We assume that $\overline{B}$ satisfies the asymmetric forcing property. We take a tournament $G = (U, F)$ and construct an undirected graph H by partitioning the vertex set uniformly at random into U_1 and U_2, such that $|U_1| = |U_2| = n/2$. We then include in H only the edges in F such that $e = (v_1, v_2)$ where $v_1 \in U_1, v_2 \in U_2$. We see that

$$t_{\overline{B}}(H) = (1 + o(1))p^{|E(B)|}$$

if and only if H is quasirandom. Now, in particular we see that $\mathbb{E}\left[\frac{|F|}{|U_1||U_2|}\right] = \frac{1}{2}$ and thus we arrive at

$$t_B(G) \geq \mu(B) + o(1).$$

To get a bound in the other direction, we embed G into a larger directed graph, D. We can construct D by uniformly at random paritioning the vertices of G as before, into U_1, U_2 such that $|U_1| = |U_2| = n/2$. We fix D such that $D - G$ is a complete, transitive, bipartite graph and U_1 and U_2 are embedded respectively into either side of the bipartite graph. We then take the underlying undirected graph, where using the fact that $\overline{B}$ satisfies the undirected forcing conjecture allows us to conclude that

$$t_B(G) \leq \mu(B) + o(1)$$

Consequently, we obtain the reverse inequality as desired:

$$t_B(G) \leq (1 + o(1))p^{|E(B)|}$$

■

We also can prove Theorem 2.4 by leveraging an intermediate lemma.

Lemma B.2. *For any fixed $\epsilon > 0$, for all nontransitive digraphs B with $|V(B)| = b > (4(1+\epsilon))^{1+1/\epsilon}$ and $|E(B)| \geq (1+\epsilon)b$, there exists a family of digraphs $(F_k)_{k=1}^{\infty}$ such that for sufficiently large k*

$$t_B(F_k) \geq (1 + \epsilon)\mu(B)t_{K_2}(F_k)^{|E(B)|}.$$

Proof. Let F_k be the digraph on kb vertices given by taking a balanced k-blowup of B (i.e. the lexicographic product of B with an empty graph on k vertices). Then,

$$t_B(F_k) \geq \frac{k^b}{(kb)^b} = b^{-b}$$

However, using the fact that $\mu(B) \leq 1$, we obtain the following

$$\begin{aligned}
\mu(B)t_{K_2}(F_k)^{|E(B)|} &\leq t_{K_2}(F_k)^{|E(B)|} \\
&\leq \left(\frac{2|E(B)|}{b^2}\right)^{|E(B)|} \\
&\leq \left(\frac{2(1+\epsilon)b}{b^2}\right)^{(1+\epsilon)b} \\
&= \left(\frac{2(1+\epsilon)}{b}\right)^{(1+\epsilon)b} \\
&= b^{-b}b^{-\epsilon b}(2+2\epsilon)^{(1+\epsilon)b} \\
&\leq \frac{1}{1+\epsilon}b^{-b}
\end{aligned}$$

where the final inequality follows by our condition that b be sufficiently large (i.e. $b > (4(1+\epsilon))^{1+1/\epsilon}$). Thus, $t_B(F_k) \geq \mu(B)t_{K_2}(F_k)^{|E(B)|}$ as desired. ■

This allows us to show the following necessary condition for a directed motif to be forcing:

Theorem. *If a digraph B that satisfies Lemma B.2 is not transitive, it is not forcing.*

Proof of Theorem 2.4. Suppose B is non-transitive, but forcing (B has at least 3 vertices). Let (G_n) be a family of transitive tournaments with G_n on n vertices, and let $G = G_n$ for some n. Then, any homomorphism $V(B) \to V(G)$ must send all edges of B to a single edge. Therefore,

$$t_B(G) = \frac{h_B(G)}{|V(G)|^{|V(B)|}} = \frac{|E(G)|}{|V(G)|^{|V(B)|}} \leq \frac{n^2}{n^{|V(B)|}} = n^{2-|V(B)|} \leq \frac{1}{n} \to 0$$

Let (F_n) be a family of digraphs on n vertices such that as $n \to \infty$

$$t_B(F_n) = (c\mu(B) + o(1))t_{K_2}(F_n)^{|E(B)|}$$

for some constant $c > 1$, constructed as in Lemma B.2. Consider the following family of digraphs (D_n) with D_n on n vertices constructed as follows. Fix a labeling of $V(G_n)$ so that $V(G_n) = \{v_1, \ldots, v_n\}$ with $\delta^+(v_i) = n - i$. Randomly label the vertices of F_n so that $V(F_n) = \{v_1, \ldots v_n\}$. Then let $V(D_n) = \{v_1, \ldots, v_n\}$ and construct $E(D_n)$ randomly as follows. Let $p = \frac{1}{c}$. Then for each i, j with $1 \leq i < j \leq n$ with probability p, construct whichever of (v_i, v_j) or (v_j, v_i) in $E(D_n)$ that is in $E(G_n)$. With probability $1-p$ do the same probabilistic edge construction using graph F_n as a reference, not constructing any edge if neither (v_i, v_j) nor $(v_j, v_i) \in E(F_n)$. Note then that the graphs D_n satisfy (as $n \to \infty$)

$$t_B(D_n) = (\mu(B) + o(1))t_{K_2}(D_n)^{|E(B)|}$$

Thus since B is forcing, the family (D_n) has quasirandom direction. However, consider the partition of $V(D_n) = V_n \sqcup W_n$ where $V_n = v_1, \ldots v_{\lfloor n/2 \rfloor}$, $W_n = v_{\lceil (n+1)/2 \rfloor}, \ldots v_n$. Note that a

$\frac{1}{c}$ random fraction of the edges are chosen for D_n in accordance with G_n and thus

$$\mathbb{E}\left[\sum_{v\in V_n}|d^+(v)-d^-(v)|\right] \geq \frac{1}{c}\sum_{i=1}^{\lfloor n/2\rfloor}(n-i)-(i-1) = \frac{1}{c}\sum_{i=1}^{\lfloor n/2\rfloor}(n-2i+1) > \sum_{j=0}^{n/5}\frac{2j}{c} \geq \frac{n}{5}\left(\frac{n}{5}-1\right) > \frac{n^2}{26c}.$$

This implies there exists some choices of a family D_n of digraphs such that for each D_n, there exists a partition of $V(D_n) = V_n \sqcup W_n$ with

$$\sum_{v\in V_n}|d^+(v)-d^-(v)| = \Omega(n^2).$$

Therefore, as in Property Q4 of [3], the family (D_n) does not have quasirandom direction, a contradiction. Consequently, B cannot be forcing. ■

Appendix C. Proofs of Theorem 2.3 and Theorem 2.4

Proof of Theorem 2.3. Fix a bipartite graph $B = (B_1 \sqcup B_2, F)$ directed such that for all $e = (b_1, b_2) \in F$, $b_1 \in B_1, b_2 \in B_2$ with underlying undirected graph $\overline{B}$ and suppose that $\overline{B}$ is forcing. Fix a tournament $T = (V, E)$ and take a random partition of the vertex set into $V = V_1 \sqcup V_2$ uniformly at random so that $|V_1| = |V_2| = n/2$. Now construct the undirected graph $H = (V, D)$ by including in D undirected edges for each $e = (v_1, v_2) \in E$ with $v_1 \in V_1$ and $v_2 \in V_2$. Since $\overline{B}$ satisfies the asymmetric Sidorenko property, we have that

$$t_{\overline{B}}(H) \geq \left(\frac{|E(H)|}{|V_1||V_2|}\right)^{|E(B)|}$$

Note that $\mathbb{E}\left[\frac{|E(H)|}{|V_1||V_2|}\right] = \frac{1}{2}$. We can count the homomorphisms of into B in G using $t_{\overline{B}}(H)$. We simply need to count the automorphisms of B as a directed graph, resulting in

$$t_B(T) \geq \mu(B)$$

■

Proof of Theorem 2.5. We consider the complete bipartite digraph on $2n$ vertices, $G = K^{\rightarrow}_{n,n} = (V_1 \sqcup V_2, E)$ where all edges $e = (v_1, v_2) \in E$ are directed so that $v_1 \in V_1$ and $v_2 \in V_2$. Let B be a non-transitive digraph (which must have at least 3 vertices and 3 edges). Suppose to the contrary that for all digraphs G, $t_B(G) \geq t_{K_2}(G)^{|E(B)|}$. Any homomorphism from B into $K^{\rightarrow}_{n,n}$ must map to a single edge. Thus, of the $|V|^{|V(B)|}$ vertex maps (considered with labels), there are at most $|E|$ homomorphisms, and thus,

$$t_B(K^{\rightarrow}_{n,n}) \leq \frac{|E|}{|V|^{|V(B)|}} = \frac{n^2}{(2n)^{|V(B)|}} = \frac{1}{2^{|V(B)|}n^{|V(B)|-2}}$$

However, note that

$$(\mu(B) + o(1))\, t_{K_2}(K^{\rightarrow}_{n,n})^{|E(B)|} \geq \frac{1}{4^{|E(B)|}}.$$

Since $|V(B)| \geq 3$, for sufficiently large n,

$$t_B(K^{\rightarrow}_{n,n}) \leq \frac{1}{2^{|V(B)|}n^{|V(B)|-2}} \ll \frac{1}{4^{|E(B)|}} \leq (\mu(B) + o(1))\, t_{K_2}(K^{\rightarrow}_{n,n})^{|E(B)|}$$

and thus B does not satisfy the desired inequality for all G, a contradiction. ■

Remark C.1. Note that we can also take G to be a large transitive tournament to show Theorem 2.5.