ON THE MAXIMUM DENSITY OF FIXED STRONGLY CONNECTED SUBTOURNAMENTS

ROBERTO F. PARENTE

Institute of Mathematics and Statistics, Universidade de São Paulo

CRISTIANE M. SATO

Center of Mathematics, Computing and Cognition, Universidade Federal do ABC

Abstract. In this paper, we study the density of fixed strongly connected subtournaments on 5 vertices in large tournaments. We determine the maximum density asymptotically for three tournaments as well as extremal families for each tournament. We use Razborov’s semidefinite method for flag algebras to prove the upper bounds.

1. Introduction

The problem of maximizing the number of strongly connected subtournaments of a fixed size in a tournament has been solved exactly by Beineke and Harary [1] in 1965. On the other hand, the related problem of minimizing transitive subtournaments of given size has only been solved recently by Coregliano and Razborov [3], who showed that the density of a transitive subtournament of any fixed size is asymptotically minimized in a tournament with \( n \) vertices when such tournament is chosen uniformly at random among all tournaments with vertex set \( [n] := \{1, \ldots, n\} \). Furthermore, they prove that, for \( k \geq 4 \), any extremal family is necessarily quasi–random.

In this paper, we study the density of fixed strongly connected subtournaments on 5 vertices. Note that Beineke and Harary [1] consider all possible strongly connected orientations, while we consider the orientations separately. We study the tournaments in Figure 1, which we denote by \( T_1, T_2, \) and \( T_3 \). We chose \( T_1 \) because of the fact that it is the only tournament on 5 vertices with

![Figure 1. Tournaments T₁, T₂ and T₃](image-url)
2 cycles of length 5. We chose $T_2$ since its maximum density is obtained by random tournaments asymptotically (and it is the only strongly connected subtournament on 5 vertices for which this holds). We chose $T_3$ since its maximum density is achieved by the family described in [1] (and this family also achieves the maximum for $T_1$).

For any tournaments $T, T'$ with $|V(T')| \leq |V(T)|$, let $p(T'; T)$ denote the probability that the tournament induced by a set of $|V(T')|$ vertices chosen uniformly at random from $V(T)$ is isomorphic to $T'$ (that is, the induced tournament is a copy of $T'$). For any positive integer $n$, let $T_n$ denote the set of tournaments with vertex set $[n]$. Next we state our main result.

**Theorem 1.1.** We have that

$$
\lim_{n \to \infty} \max_{T_n \in T_n} \{p(T_1; T_n)\} = \frac{1}{16}, \tag{1}
$$

$$
\lim_{n \to \infty} \max_{T_n \in T_n} \{p(T_2; T_n)\} = \frac{15}{128}, \tag{2}
$$

and

$$
\lim_{n \to \infty} \max_{T_n \in T_n} \{p(T_3; T_n)\} = \frac{5}{16}. \tag{3}
$$

Equation (1) partially confirms a conjecture proposed by Coregliano in [2].

We split the proof of Theorem 1.1 into two parts: the proof of the lower bounds and the upper bounds. The proof of the upper bounds are given in Section 3 and are an application of Razborov’s semidefinite method for flag algebras [6]. In Section 2 we present a brief overview of this method. The proofs of the lower bounds are given in Section 4 where we show families of tournaments $(R_n)_{n \in \mathbb{N}}$ achieving the values in (1), (2) and (3).

2. RAZBOROV’S SEMIDEFINITE METHOD FOR FLAG ALGEBRAS

In this section we present a brief summary of the semidefinite method proposed by Razborov [6]. We state the method for tournaments, but we mention the method works for the very general setting of universal theories.

Given a tournament $H$, for every integer $\ell$, let $\mathcal{H}_\ell$ be the set of tournaments on $\ell$ vertices that are subtournaments of $H$ up to isomorphism. Then we can compute $p(T', T)$ as an average over $\mathcal{H}_\ell$ as follows

$$
p(T'; T) = \sum_{H \in \mathcal{H}_\ell} p(T'; H)p(H; T) \leq \max_{H \in \mathcal{H}_\ell} \{p(T'; H)\}. \tag{4}
$$

However, in general, this upper bound is too weak to help finding the extremal values. Next we will see how to use flag algebras to obtain better bounds. We start by defining some basics notions of flag algebras.

Let $T$ be a $n$-vertex tournament, $k$ a positive integer and $\theta$ an injective map $\theta: [k] \to V(T)$. A flag is the pair $F = (T, \theta)$. Given a flag $F$, we denote by $T_F$ the tournament associated with the flag. If $\theta$ is bijective we call the flag a type.

Given a flag $F = (T_F, \theta)$, let $T_F[\theta]$ denote the tournament of $[k]$ such that $(i, j)$ is an arc in $T_F[\theta]$ if $(\theta(i), \theta(j))$ is an arc in $T_F$. Given a type $\sigma = (T_\sigma, \theta_\sigma)$, we say that a flag $F = (T_F, \theta)$ is $\sigma$-flag if $T_F[\theta] = T_{\sigma}[\theta_\sigma]$. The order of $\sigma$-flag $F$ is $|V(T_F)|$. We say that $\sigma$-flags $F = (T_F, \theta)$ and $F' = (T'_{F'}, \theta')$ are isomorphic if there exists an (digraph) isomorphism $\phi$ for $F$ to $F'$ such that $\phi(\theta(i)) = \theta'(i)$ for all $i \in [k]$. 

2
Fix a type $\sigma$ and an integer $m$. Let $F^\sigma_m$ be the set of all $\sigma$-flags of order $m$, up to isomorphism.

Now we define a probability involving the embedding two $\sigma$-flags in another $\sigma$-flag so that the labeled vertices coincide. Let $F_a$ and $F_b$ be $\sigma$-flags of order $m_a$ and $m_b$, respectively, and let $F = (T_F, \theta)$ be a $\sigma$-flag. Let $V_a$ be a $m_a$-set and $V_b$ be a $m_b$-set chosen uniformly at random from $V(T_F)$ such that $V_a \cap V_b = \text{im}(\theta)$. Define $p(F_a, F_b; \theta; T_F)$ as the probability that $(T_F[V_a], \theta)$ is isomorphic to $F_a$ and $(T_F[V_b], \theta)$ is isomorphic to $F_b$. Furthermore, define $p(F_a, \theta; T_F) = p(F_a, \sigma, \theta; T_F)$.

Note that the difference between $p(F_a, \theta; T_F)p(F_b, \theta; T_F)$ and $p(F_a, F_b; \theta; T_F)$ comes from sampling of $V_a$ and $V_b$ with or without replacement. The following lemma tells us that this difference is essentially negligible (this is a special case of Lemma 2.3 in [5]).

**Lemma 2.1** (Razborov [5]). For any $F_a, F_b \in F^\sigma_m$, and $\theta \in \Theta$, 
\[
p(F_a, \theta; T_F)p(F_b, \theta; T_F) = p(F_a, F_b, \theta; T_F) + o(1),
\]
where the $o(1)$ tends to zero as $|V(T_F)|$ tends to infinity.

For all $k$ and tournament $T$, let $\Theta(k, T)$ be the set of all injective functions from $[k]$ to $V(T)$. Let $F$ be a $\sigma$-flag and $\theta$ be chosen uniformly at random from $\Theta = \Theta(|\sigma|, T_F)$. By linearity of expectation and Lemma 2.1, we have 
\[
E_{\theta \in \Theta}[p(F_a, \theta; T_F)p(F_b, \theta; T_F)] = E_{\theta \in \Theta}[p(F_a, F_b, \theta; T_F)] + o(1)
\]
(5)

For every integer $k$ and subtournament $H$ of a tournament $T$, let $\Theta_H = \Theta_H(k)$ be the restriction of $\Theta(k, T)$ to the functions $\theta$ such that im($\theta$) $\subseteq V(H)$. Now, we compute the expectation of $p(F_a, F_b, \theta; T)$ by first embedding $F_a$ and $F_b$ into subtournaments of $T$. We have that 
\[
E_{\theta \in \Theta}[p(F_a, F_b, \theta; T)] = \sum_{H \in \mathcal{H}_\ell} E_{\theta \in \Theta_H}[p(F_a, F_b, \theta; H)]p(H; T).
\]
(6)

For every $\theta \in \Theta = \Theta(|\sigma|, T)$ define the vector $p_\theta = (p(F, \theta; T) : F \in F^\sigma_m)$ indexed by $F^\sigma_m$.

We have all necessary definitions to present the semidefinite method and improve the upper bound in [4]. For this we use properties of positive semidefinite matrices. Let $Q$ be a positive semidefinite matrix indexed by $F^\sigma_m \times F^\sigma_m$. By the definition of positive semidefinite matrices, we have that $p^TQp \geq 0$ for all real vectors $p$ indexed by $F^\sigma_m$. Thus, by applying this inequality to $p_\theta$ for $\theta \in \Theta(|\sigma|, T)$ and taking the average over $\Theta$, we conclude $E_{\theta \in \Theta}[p_\theta^TQp_\theta] \geq 0$. Using (5), (6) and by linearity of expectation, we have 
\[
E_{\theta \in \Theta}[p_\theta^TQp_\theta] = \sum_{H \in \mathcal{H}_\ell} \left( \sum_{F_a, F_b \in F^\sigma_m} Q_{F_a, F_b} E_{\theta \in \Theta_H}[p(F_a, F_b, \theta; H)] \right) c_H(\sigma, m, Q) p(H; T) + o(1),
\]
(7)
where we define $c_H(\sigma, m, Q)$ as the coefficient of $p(H; T)$ for each $H \in \mathcal{H}_\ell$ in the summation above.

Given a positive integer $t$, let $(\sigma_i, m_i, Q_i)$ be triples such that, for every $1 \leq i \leq t$, we have that $\sigma_i$ is a type, $m_i \leq (\ell + |\sigma_i|)/2$ is an integer, and $Q_i$ is a positive semidefinite matrix of
dimension $|\mathcal{F}_{m_i}^\sigma|$. For $H \in \mathcal{H}_\ell$, define
\[
c_H = \sum_{i=1}^t c_H(\sigma_i, m_i, Q_i).
\]
(8)

Note that $c_H$ can be negative. Because of this, it can be used to improve the upper bound in (4). Since $E_{\theta \in \Theta} [p_T^\theta Q_{\theta}] \geq 0$, by (4),
\[
p(T'; T) \leq E_{\theta \in \Theta} [p_T^\theta Q_{\theta}] + \sum_{H \in \mathcal{H}_\ell} p(T'; H)p(H; T) = \sum_{H \in \mathcal{H}_\ell} (c_H + p(T'; H))p(H; T)
\]
(9)
\[
\leq \max_{H \in \mathcal{H}_\ell} \{c_H + p(T'; H)\}.
\]
The value of $c_H$ is computed by the following semidefinite program
\[
\begin{align*}
\text{minimize} & \quad Y \\
\text{subject to} & \quad c_H + p(T'; H) \leq Y, \quad \forall H \in \mathcal{H}_\ell, \\
& \quad Q \succeq 0.
\end{align*}
\]
(10)

3. Proof of Theorem 1.1: upper bounds

In this section we prove the upper bounds in Theorem 1.1. We use the semidefinite method for flag algebras as presented in Section 2.

Lemma 3.1. For every $n$-vertex tournament $T_n$,
\[
\lim_{n \to \infty} p(T_1; T_n) \leq \frac{1}{16}.
\]
(11)

Proof of Lemma 3.1. In order to use the semidefinite method, we need to fix $\ell$, which is used in [4]. Then we need to define $c_H$ for every $H \in \mathcal{H}_\ell$ as in (8). To define $c_H$ we choose how many
types $t$ we will use and the types $\sigma_i$ we want to use. For each type $\sigma_i$, we choose an integer $m_i$ satisfying $m_i \leq (\ell + |\sigma_i|)/2$ and a positive semidefinite $|\mathcal{F}_{m_i}^\sigma| \times |\mathcal{F}_{m_i}^\sigma|$ matrix $Q_i$.

We will use $t = 2$ and use the types $\sigma(3, 1)$ and $\sigma(3, 2)$ defined as follows. Let $T(3, 1)$ and $T(3, 2)$ be the two 3-vertex non-isomorphic tournaments, and let $\theta_1: [3] \to V(T_3^1)$ and $\theta_2: [3] \to V(T(3, 2))$ be two injective mapping such that $\sigma(3, 1) = (T(3, 1), \theta_1)$ and $\sigma(3, 2) = (T(3, 2), \theta_2)$ are types (see Figure 2). Define $\sigma_1 = \sigma(3, 1)$ and $\sigma_2 = \sigma(3, 2)$.

\[\begin{array}{c}
\begin{array}{c}
1 \\
2 \\
3
\end{array} \\
\begin{array}{c}
1 \\
2 \\
3
\end{array}
\end{array}\]

(a) $\sigma(3, 1)$

(b) $\sigma(3, 2)$

Figure 2. Types $\sigma(3, 1)$ and $\sigma(3, 2)$
We use $\ell = 6$ and $m_1 = m_2 = 4$ (so that $4 = m_i \leq (\ell + |\sigma_i|)/2 = 4.5$). We use the positive semidefinite matrices $Q(1, 1)$ and $Q(1, 2)$ with dimension $|F^\sigma_{4_1^m}| = |F^\sigma_{4_2^m}| = 8$ shown below:

\[
Q_1 := Q(1, 1) = \frac{1}{8}
\begin{pmatrix}
11 & 15 & 15 & -15 & 15 & -15 & -15 & -11 \\
15 & 25 & 25 & -25 & 25 & -25 & -25 & -15 \\
15 & 25 & 25 & -25 & 25 & -25 & -25 & -15 \\
-15 & -25 & -25 & 25 & -25 & 25 & 25 & 15 \\
15 & 25 & 25 & -25 & 25 & -25 & -25 & -15 \\
-15 & -25 & -25 & 25 & -25 & 25 & 25 & 15 \\
-15 & -25 & -25 & 25 & -25 & 25 & 25 & 15 \\
-11 & -15 & -15 & 15 & -15 & 15 & 15 & 11
\end{pmatrix}
\]  

\[Q_2 := Q(1, 2) = \frac{1}{8}
\begin{pmatrix}
15 & -15 & -12 & 15 & -15 & 12 & 15 & -15 \\
-15 & 15 & 12 & -15 & 15 & -12 & -15 & 15 \\
-12 & 12 & 15 & -12 & 12 & -15 & -12 & 12 \\
15 & -15 & -12 & 15 & -15 & 12 & 15 & -15 \\
-15 & 15 & 12 & -15 & 15 & -12 & -15 & 15 \\
12 & -12 & -15 & 12 & -12 & 15 & 12 & -12 \\
15 & -15 & -12 & 15 & -15 & 12 & 15 & -15 \\
-15 & 15 & 12 & -15 & 15 & -12 & -15 & 15
\end{pmatrix}.
\]

To see that $Q_1$ and $Q_2$ are positive semidefinite, we analyse their characteristic polynomials: $p_{Q_1}(x) = x^8 - 43/2x^7 + 75/8x^6$ and $p_{Q_2}(x) = x^8 - 15x^7 + 243/16x^6$. It is easy to see that these polynomials have non-negative roots and so the matrices $Q_1$ and $Q_2$ are positive semidefinite. The matrices $Q_1$ and $Q_2$ were found with the aid of semidefinite programming solver \cite{4} for the SDP in \cite{10}.

We compute $c_H = c_H(\sigma_1, 4, Q_1) + c_H(\sigma_2, 4, Q_2)$ and $p(\cdot; H)$, for every $H \in \mathcal{H}_6$, and obtain

\[
\lim_{n \to \infty} p(T_1; T_n) \leq \max_{H \in \mathcal{H}_6} \{ p(T_1; H) + c_H \} = 1/16.
\]

The set $\mathcal{H}_6$ consists of 56 non-isomorphic tournaments and we compute $p(T_1; H)$ for each $H \in \mathcal{H}_6$ exactly using computational tools.

The proofs of the upper bounds for $T_2$ and $T_3$ are very similar to the proof of Lemma \cite{3}. We choose an integer $m_i$ satisfying $m_i \leq (\ell + |\sigma_i|)/2$ and a positive semidefinite $|F^\sigma_{m_i}| \times |F^\sigma_{m_i}|$ matrix $Q_i$.

In the tables below, $\sigma(3, 1)$ and $\sigma(3, 2)$ are as defined in the proof of Lemma \cite{3} and $\sigma(1)$ consists of a tournament with a single vertex labeled by 1. The matrices are shown in Appendix \ref{A}. As in the proof of Lemma \cite{3} to show that these are positive semidefinite, we compute their characteristic polynomials using Maxima (a symbolic mathematics software) and see that their roots are non-negative.

For each $j \in \{2, 3\}$, we then compute $c_H = \sum_{i=1}^{t} c_H(\sigma_i, m_i, Q_i)$ and $p(T_j; H)$, for every $H \in \mathcal{H}_t$, and obtain the desired bounds.
4. Proof of Theorem 1.1 Lower Bound

In this section, we prove the lower bounds in Theorem 1.1.

Lemma 4.1. There exists a family of tournaments \( (R_n)_{n \in \mathbb{N}} \) such that \( R_n \) has \( n \) vertices for each \( n \) and satisfies

\[
\lim_{n \to \infty} p(T_1; R_n) = \frac{1}{16}.
\]

and

\[
\lim_{n \to \infty} p(T_3; R_n) = \frac{5}{16}.
\]

Proof of Lemma 4.1. We omit the proof for \( T_3 \) since it is very similar to the proof for \( T_1 \). Fix \( n > 5 \) and \( R_n \) is the \( n \)-tournament constructed as follows. Let \( R_n \) be a tournament on \([n]\) and, for each distinct \( i, j \in [n] \), include the arc \((i, j)\) if \( j - i + 1 \leq n/2 \) and the arc \((j, i)\) if \( j - i + 1 > n/2 \). Now, we compute the number of (induced) copies of \( T_1 \) in \( R_n \), which we denote by \( \text{ind}(T_1; R_n) \). Suppose we label the vertices of \( T_1 \) from 1 up to 5.

Suppose that the vertex 1 from \( T_1 \) is mapped to the vertex 1 of \( R_n \). If vertex 2 is mapped to a vertex \( i \), then \( 2 \leq i \leq \lfloor n/2 \rfloor \) and vertex 3 has to be mapped to a vertex \( j \) such that \( i + 1 \leq j \leq \lfloor n/2 \rfloor \). Vertex 4 has to be mapped to a vertex \( k \) such that \( \lfloor n/2 \rfloor + 1 \leq k \leq \lfloor n/2 \rfloor - 1 + i \) (since it \((2, 4)\) and \((4, 1)\) are arcs in \( T_1 \)). Vertex 5 has to be mapped to a vertex \( \ell \) such that \( \lfloor n/2 \rfloor + i \leq \ell \leq \lfloor n/2 \rfloor - 1 + j \) (since it \((3, 5)\) and \((5, 2)\) are arcs in \( T_1 \)). See Figure 3.

![Figure 3. Possibilities of embedding \( T_1 \) in \( R_n \)](image_url)

Note that, after we fix the positions of the vertices 1, 2 and 3, the number of choices for the vertex 4 becomes \( i - 1 \) and for the vertex 5 becomes \( j - i \). A similar analysis for each of the \( n \)
choices for vertex 1 shows that, after we fix the positions of the vertices 1, 2 and 3, the number of choices for the vertex 4 becomes \(i + O(1)\) and for the vertex 5 becomes \(j - i + O(1)\). Thus, we have \(n\) choices for vertex 1 and \(\sum_{i=1}^{\lfloor n/2 \rfloor} \sum_{j=i+1}^{\lfloor n/2 \rfloor} (i + O(1))(j - i + O(1))\) for the other vertices. We divide this number by 5 to account for the number of automorphisms of \(T_1\). Hence, we have that

\[
\text{ind}(T_1; R_n) = \frac{n}{5} \cdot \sum_{i=1}^{\lfloor n/2 \rfloor} \sum_{j=i+1}^{\lfloor n/2 \rfloor} (i + O(1))(j - i + O(1))
\]

\[
= \frac{n}{5} \cdot \sum_{i=1}^{\lfloor n/2 \rfloor} \sum_{j=i}^{\lfloor n/2 \rfloor} i(j - i) + O(n^4)
\]

\[
= \frac{n}{5} \cdot \left( \frac{1}{2} \sum_{i=1}^{\lfloor n/2 \rfloor} i(\lfloor n/2 \rfloor - i + 1)(\lfloor n/2 \rfloor + i) - \sum_{i=1}^{\lfloor n/2 \rfloor} i^2(\lfloor n/2 \rfloor - i + 1) \right) + O(n^4)
\]

\[
= \frac{n}{5} \cdot \left( \frac{1}{2} \sum_{i=1}^{\lfloor n/2 \rfloor} i + 1 \sum_{i=1}^{\lfloor n/2 \rfloor} i^3 - \frac{\lfloor n/2 \rfloor}{2} \sum_{i=1}^{\lfloor n/2 \rfloor} i^2 \right) + O(n^4)
\]

\[
= \frac{n^5}{5 \cdot 2^4} \left( \frac{1}{4} + \frac{1}{8} - \frac{1}{3} \right) + O(n^4) = \frac{n^5}{5 \cdot 2^7 \cdot 3} + O(n^4).
\]

and so

\[
\lim_{n \to \infty} p(T_1; R_n) = \lim_{n \to \infty} \frac{\text{ind}(T_1; R_n)}{\binom{n}{5}} = \frac{5!}{2^7 \cdot 3 \cdot 5} = \frac{1}{16}. 
\]

\(\square\)

Let \(T(n)\) denote the tournament on \([n]\) obtained by choosing a orientation of the complete graph on \([n]\) uniformly at random. Since the probability that the tournament induced by any 5 vertices of \(T(n)\) is isomorphic to \(T_2\) is \((1/2)^{10} \cdot 5!/|\text{Aut}(T_2)|\), we have that

\[
\lim_{n \to \infty} \mathbb{E}(p(T_2; T(n))) = \frac{15}{128},
\]

which proves the lower bound for \(T_2\) in Theorem 1.1.

---

**Appendix A.**

\[
Q(2, 1) = \begin{pmatrix}
0.3745878162 & -0.4878738581 & -0.2289130375 & 0.3421990795 \\
-0.4878738581 & 0.7596348467 & 0.2161092295 & -0.4878738581 \\
-0.2289130375 & 0.2161092295 & 0.241768455 & -0.2289130375 \\
0.3421990795 & -0.4878738581 & -0.2289130375 & 0.3745878162
\end{pmatrix}
\]
The maximum number of strongly connected subtournaments

L. Nagami Coregliano and A. A. Razborov, Quasi-Carousel Tournaments

L. Nagami Coregliano, Flag algebras

Alexander Razborov, Flag algebras, J. Symbolic Logic 72 (2007), no. 4, 1239–1282. MR 2371204 (2008j:03040)

On 3-hypergraphs with forbidden 4-vertex configurations, SIAM J. Discrete Math. 24 (2010), no. 3, 946–963. MR 2680226 (2011k:05171)