Discrete Applied Mathematics 176 (2014) 30–42

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Discrepancy inequalities for directed graphs

Fan Chung [a], Franklin Kenter [b,*]

[a] University of California, San Diego, La Jolla, CA 92093, United States
[b] Rice University, Houston, TX 77005, United States

ARTICLE INFO

Article history:
Received 30 August 2012
Received in revised form 4 March 2014
Accepted 30 March 2014
Available online 21 April 2014

Keywords:
Directed graph
Discrepancy
Eigenvalues
Singular values
Skew-symmetric matrix

ABSTRACT

We establish several discrepancy and isoperimetric inequalities for directed graphs by considering the associated random walk. We show that various isoperimetric parameters, as measured by the stationary distribution of the random walks, including the Cheeger constant and discrepancy, are related to the singular values of the normalized probability matrix and the normalized Laplacian. Further, we consider the *skew-discrepancy* of directed graphs which measures the difference of flow among two subsets. We show that the skew-discrepancy is intimately related to $\mathcal{Z}$, the skew-symmetric part of the normalized probability transition matrix. In particular, we prove that the skew-discrepancy is within a logarithmic factor of $\|\mathcal{Z}\|$. Finally, we apply our results to construct extremal families of directed graphs with large differences between the discrepancy of the underlying graph and the skew-discrepancy.

1. Introduction

The discrepancy of a graph bounds the largest difference between the number of edges between two subsets of vertices and its expected value among all possible choices of subsets. The study of discrepancy in graph theory has been an extensively useful tool in spectral graph theory with wide applications in extremal graph theory, in the analysis of approximation algorithms, and in statistical tests (see [10]).

In the undirected case, there are several different ways to measure the discrepancy. One way is to consider the size of the two subsets, and take the "expected" number of edges to be proportional to the their product. This is a natural consideration for regular graphs, as in this case, the discrepancy of $A, B \subset V$ is bounded above by $\lambda_2\sqrt{|A||B|}$ where λ_2 is the second largest absolute eigenvalue of the adjacency matrix of G [2]. For a graph with a general degree sequence, the normal eigenvalue bound using the adjacency matrix does not apply. Instead, one may consider the *volume*, the sum of the degrees, of each subset; in which case, the "expectation" is taken to be proportional to the product of the volumes. Under this notion, the discrepancy can be bounded using the spectral gap of the *normalized Laplacian matrix* (see [7]).

This study focuses on directed graphs by adapting the concepts above. However, this presents additional challenges. First, directed graphs do not have a natural notion of degree or volume, as the in- and out-degree of each vertex do not necessarily coincide. Further, many of the graph theoretic matrices of directed graphs are not symmetric; therefore, many of the tools and techniques used to study undirected graphs cannot be applied to the general directed case.

To address the first challenge, in their work on the Cheeger constant, Chung [8] followed by Li and Zhang [11] consider a *random walk* and apply the stationary distribution in order to measure the *volume* of a subset of vertices instead of size of the sets. Specifically, in our case, we use the stationary distribution of a typical random walk on a directed graph G to

* Corresponding author. Tel.: +1 650 465 2245; fax: +1 858 534 5273.
E-mail addresses: fan@ucsd.edu (F. Chung), franklin.kenter@gmail.com, franklin.h.kenter@rice.edu (F. Kenter).

http://dx.doi.org/10.1016/j.dam.2014.03.023
0166-218X/

define two types of discrepancies: $\mathrm{disc}(G)$ and $\mathrm{disc}'(G)$. Roughly speaking, $\mathrm{disc}(G)$ bounds the difference of the flow from one subset of vertices S to another subset T from the expected quantity while $\mathrm{disc}'(G)$ measures the difference between the flow from S to T and the reverse flow from T to S. Note that the expected quantity depends on the stationary distribution and can be quite different in one direction from the other. Precise definitions are given in Section 2.

To overcome the second obstacle, one can symmetrize the matrix associated with the directed graph (or the associated random walk) as in [1,8], or [11]. As a result, one can apply the techniques used with symmetric matrices. In doing so, one loses information regarding the directed nature of the graph. In order to capture the directed nature of a directed graph, Li and Zhang in [11] considered the skew-symmetric part of the normalized probability transition matrix.

In this paper, we expand upon this method of using three graph theoretic matrices and show their eigen- and singular values bound the notions of discrepancy described above. We will show that these two types of discrepancies are intimately related to several types of eigenvalues. The first type is derived using the (normalized) Laplacian $\mathcal{L}$ as defined earlier in [8] to establish a generalized Cheeger's inequality for directed graphs. For undirected graphs, the eigenvalues of the Laplacian $\mathcal{L}$ can be used to bound the discrepancy. Here, we can still use eigenvalues of $\mathcal{L}$ to bound the discrepancy with regard to the flow from a subset S to its complement, $\bar{S}$. To deal with the discrepancy from a subset S to another subset T in a directed graph G, we will use singular values of the normalized transition probability matrix, $\mathcal{P}$, of the random walk on G. Another type of eigenvalue depends on the skew-symmetric matrix $\mathcal{Z}$ that will be defined in Section 2 and is useful for bounding $\mathrm{disc}'(G)$.

The paper is organized as follows: In Section 2, we give basic definitions. In Section 3, we derive facts regarding the graph theoretic matrices we use. In Section 4, we deal with the discrepancy between a subset and its complement and derive the relation with eigenvalues of the Laplacian $\mathcal{L}$. We then bound the discrepancy for any two general subsets in terms of the singular values of the transition probability matrix in Section 5. In Sections 6 and 7, we give several upper bounds for the skew-discrepancy, $\mathrm{disc}'(G)$, in terms of the matrix $\mathcal{Z}$ and its maximal singular value, $\|\mathcal{Z}\|$. Finally, we give constructions and applications of these results in Section 8.

2. Preliminaries

For a directed graph $G = (V, E)$ with edge weights $w_{u,v} > 0$ we consider the associated random walk on G whose transition probability matrix, denoted $\mathbf{P}$, is given by:

$$\mathbf{P}(u, v) = \frac{w_{u,v}}{d_u}$$

where $d_u = \sum_v w_{u,v}$ denotes the total weights among out-going arcs of u.

We say a directed graph is *aperiodic* if the greatest common divisor of the lengths of all closed walks is 1. Otherwise, it is *periodic*. A random walk is said to be *ergodic* if its directed graph is strongly connected and aperiodic [1]. An ergodic random walk has a unique stationary distribution, ϕ, obeying

$$\boldsymbol{\phi}\mathbf{P} = \phi. \tag{1}$$

For the purposes of this paper, in order to guarantee a unique stationary distribution, we only consider directed graphs which are aperiodic and strongly connected.

Note that ϕ can be used to define a special type of flow on the edges of G, called *circulation* as follows (see [8]). For an edge (u, v), the flow $f_\phi(u, v)$ is

$$f_\phi(u, v) = \phi(u)\mathbf{P}(u, v).$$

For two subsets S and T of vertices in G the flow from S to T is denoted by:

$$f(S, T) = \sum_{u\in S, v\in T} f(u, v).$$

We extend the notion of ϕ to subsets of vertices by defining

$$\phi(S) = \sum_{v\in S} \phi(v).$$

For two subsets of vertices S, T, we define $\mathrm{disc}(S, T)$ to be the quantity

$$\mathrm{disc}(S, T) = |f_\phi(S, T) - \phi(S)\phi(T)|.$$

Notice that for an undirected graph, $\phi(v) = \frac{d_v}{\mathrm{vol}\, G}$, where $\mathrm{vol}\, G = \sum_v d_v$ and therefore $f_\phi(u, v) = 1/\mathrm{vol}\, G$. Hence, the above notion of discrepancy is consistent with that for undirected graphs using the normalized Laplacian as seen in [7].

Further, we define the *skew-discrepancy*, denoted $\mathrm{disc}'(S, T)$, to be the quantity

$$\mathrm{disc}'(S, T) = \left|\frac{f_\phi(S, T) - f_\phi(T, S)}{2}\right|.$$