

Multicommodity Allocation for Dynamic Demands Using PageRank Vectors

Fan Chung, Paul Horn, and Jacob Hughes

Abstract. We consider a variant of the contact process concerning multicommodity allocation. In this process, the demands for several types of commodities are initially given at some specified vertices, and then the demands spread interactively in the contact graph. To allocate supplies in such a dynamic setting, we use a modified version of PageRank vectors, called Kronecker PageRank, to identify vertices for shipping supplies. We analyze both the situation that the demand distribution evolves mostly in clusters around the initial vertices and the case that the demands spread to the whole network. We establish sharp upper bounds for the probability that the demands are satisfied as a function of PageRank vectors.

I. Introduction

Efficient allocation of resources to meet changing demands is a task arising in numerous applications. For example, institutions such as governments and corporations respond to the needs of a populace, and wish to meet the demands within

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uinm.

a limited expenditure of resources. In some cases in which demand spreads, one has to be able to act before demand becomes unmanageable. In the case of an epidemic, for instance, it is desirable to find a way to distribute medicine so that the disease can be contained. Such problems have been studied in several contexts using the contact process model [Kiss et al. 06, Ganesh et al. 05, Borgs et al. 10, Newman 02, Chung et al. 09]. In [Chung et al. 09], it was demonstrated how to use PageRank vectors both to restrict the number of nodes inoculated and to provide certain containment guarantees.

In this paper, we study a variant of the classical contact process, a continuous-time Markov process on a contact graph. This model was previously used for modeling the spread of disease. In our scenario, vertices in the graph each have varying levels of demand for multiple commodities. Demand at a vertex propagates to its neighbors at a rate depending on the current demand. Our model allows for rich interactions between different commodities; for instance, demand for one commodity can influence demand for another. This fits many scenarios that arise. For example, the demand for iPhones may surpass the demand for iPads.

As another example, demand at a node can sometimes be viewed as a measure of discontent with the current supply of a resource. It is natural for an unhappy node to create unrest in its neighbors. As the contact process continues, the demands at a vertex are increased based on the demands at neighboring vertices and are decreased at a satisfaction rate, which can be thought of as a frequency of shipments. The rates at which demand spreads will be a linear combination of demands from neighboring vertices. These rates will be encapsulated as a spread matrix B , roughly analogous to the infectivity parameter in the classical contact process. The goal of this paper is to find satisfaction rates depending on the spread matrix B and the geometry of the contact graph that ensure that all vertices have no demand and the process dies out. This process will be defined, in detail, in Section 2.

To satisfy the demands that evolve according to our model as defined in Section 2, the goal is to ship commodities and supply to vertices with unsatisfied demands in an efficient way. The model here differs somewhat from typical resource-allocation problems in the sense that we do not specify the location of the “warehouses” for the supply. We will be concerned with neither the sources of the supply nor the detailed incremental costs of shipping. Instead, our goal is to identify how often to ship each commodity to a particular vertex in order to contain and satisfy demand, given an initial seed set. The reader is referred to [Chevalyre et al. 06] for the classical resource-allocation problem.

After we describe the demand model in Section 2, we proceed to analyze our supply scheme. First, we introduce the Kronecker PageRank in Section 3 based on the PageRank originally introduced in [Brin and Page 98]. Our analysis

consists of two parts. First, we give conditions that ensure that all demand is satisfied in $O(\log n)$ time with high probability, regardless of the initial demand. This is a global solution in the sense that it involves “scheduling shipments” to all vertices in the graph in a way that will be made precise once the model is formally introduced in Section 2 and Theorem 4.1. Next, in Section 5, we analyze a situation whereby shipments are scheduled to only a subset of vertices containing the initial demand. In particular, when the contact graph has some clustering structure, we are interested in subsets such that the demand within the subset is satisfied quickly (in $O(\log n)$ time) and demands reach a vertex not receiving shipments with low probability. Precise results in this direction are given in Section 5.

Our analysis provides a tradeoff in the following sense: On the one hand, if one would like to guarantee that the demand escapes a set with probability at most ϵ , our results will allow us to use PageRank (or a sharper Kronecker PageRank as defined in Section 3) to identify a subset of vertices to which supplies will be shipped and provide a guarantee that this will be sufficient. On the other hand, if one would like to send shipments to a particular set of vertices, then our analysis in Theorem 5.3 allows a guaranteed bound on the escape probability that depends on the clustering structure of the contact graph.

2. The Demand Model

We model the demand spreading along an undirected simple graph $G = (V, E)$. We write this $v \sim w$ when v and w are adjacent. For each vertex $v \in V$, let d_v be the degree of v , which is the number of neighbors of v . We adopt the convention for this discussion that each vertex is adjacent to itself, and d_v includes v in the count of neighbors. We let $n = |V|$, the number of nodes of G . Let D be the diagonal degree matrix with entries $D_{vv} = d_v$, and let A be the adjacency matrix with entries

$$A_{vw} = \begin{cases} 1, & \text{if } v \sim w, \\ 0, & \text{otherwise.} \end{cases}$$

A walk is a sequence of vertices (v_0, v_1, \dots, v_k) , where $v_i \sim v_{i+1}$. A random walk of length k is a sequence of random variables (x_0, \dots, x_k) , where the starting vertex x_0 is chosen according to some distribution, and

$$\mathbb{P}(x_{i+1} = v \mid x_i) = \begin{cases} 1/d_{x_i}, & \text{if } x_i \sim v, \\ 0 & \text{otherwise.} \end{cases}$$

The transition probability matrix for a random walk on G is given by $W = D^{-1}A$.

If B is a $k \times k$ matrix and A an $n \times n$ matrix, then the Kronecker product $A \otimes B$ is the $nk \times nk$ block matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{pmatrix}.$$

An exponential random variable with parameter λ has probability density function given by $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$, and 0 for $x < 0$. This distribution will be denoted by $\text{Exp}(\lambda)$. Exponential random variables are memoryless, that is, if X is an exponential random variable, then for constants $a, b > 0$, one has

$$\mathbb{P}(X > a + b \mid X > a) = \mathbb{P}(X > b).$$

If X and Y are independent and $X \sim \text{Exp}(\lambda_1)$, $Y \sim \text{Exp}(\lambda_2)$, then $\min\{X, Y\} \sim \text{Exp}(\lambda_1 + \lambda_2)$. A Poisson point process at rate λ is a sequence of random variables X_1, X_2, \dots such that $X_i - X_{i-1}$ has distribution $\text{Exp}(\lambda)$.

Before we describe our model, let us briefly recall the notion of a contact process on a graph G , which we denote by $\text{CP}(T, \beta, \sigma, G)$. In a contact process (see, for example, [Borgs et al. 10] or [Chung et al. 09]), a disease initially infects a set $T \subseteq V(G)$. The disease has an infectivity parameter β , and each vertex has a certain amount of “medicine” σ_v . Each infected vertex independently infects its neighbors at times given by a Poisson point process at rate β , and each infected vertex is cured at times given by a Poisson point process at rate σ_v . In the most frequently studied case, σ is constant and the host graph is an infinite graph. The process ends when all vertices are cured, and the basic problem is to determine under what conditions on σ and β the process ends almost surely (a.s.). In the case of finite graphs, if $\sigma_v > 0$ for every vertex, it is easy to observe that the process ends a.s., so the problem becomes determining how fast the process ends.

The k -commodity dynamic demand model on a graph G is a variant of the contact process, $\text{DD}(\tau(0), B, \sigma, G)$. In this situation, the spread matrix B is a $k \times k$ real matrix (not assumed to be symmetric or even nonnegative) along with a supply function $\sigma : V \rightarrow \mathbb{R}^k$ and the initial demand $\tau(0) : V \rightarrow \mathbb{N}^k$. At time t , each node v has demand $\tau_v(t) \in \mathbb{N}^k$, with each coordinate representing a different product in demand. A node v is said to be *satisfied* at time t if $\tau_v(t) = \mathbf{0}$, and *unsatisfied* otherwise. The state of the process is described by the demand vector τ , where $\tau_v^j(t)$ is the demand for commodity j at node v at time t .

The spread matrix $B = [\beta_{ij}]$ describes how the demand for one commodity influences demands for other commodities. The i, j entry of B , β_{ij} , determines the spread rate of the demand for commodity j that is caused by demand for

commodity i . In particular, we can describe the rate of spread events as follows. If v is a node that is unsatisfied at time t , and w an adjacent vertex, then there are spread events from v to w with rates $\max\{\tau_v(t)B, 0\}$. That is, the rate at which τ_w^j increases due to the demand at v is given by $\max\{\sum_i \tau_v^i(t)\beta_{ij}, 0\}$. Here, when we say that an event occurs with rate λ , we mean that the elapsed time until that event takes place is distributed as $\text{Exp}(\lambda)$. Because the minimum of exponential random variables is itself an exponential random variable, we can capture the total spreading rates in a condensed form. We define the rate function $\rho(t) : V \rightarrow \mathbb{R}^k$ at time t by

$$\rho_v = \sum_{w \sim v} \tau_w(t)B = (\tau(t)(A \otimes B))_v,$$

where $\tau(t)$ is viewed as a vector with indices indexed by $V \times k$, and $\rho_v^i(t)$ is the rate at which τ_v^i is increasing at time t .

Supply events occur with rates given by $\tau(t)\text{Diag}(\sigma)$, independently of any neighboring supply events. That is, the time until τ_v^i is decreased by 1 is distributed as $\text{Exp}(\sigma_v^i \tau_v^i)$.

We briefly give a construction of the process to show that it is well defined. Let \vec{E} denote the set of ordered edges, that is, ordered pairs that are edges in the graph, so that uv and vu are distinct. We run independent Poisson point processes

$$\{X_e^{j,\rho}\}_{e \in \vec{E}(G), j \in [k], \rho \in \mathbb{N}^k},$$

so that $X_e^{j,\rho}$ is at rate $\max\{0, [\tau_v B]_j\}$, and independent Poisson point processes

$$\{X_v^{i,n}\}_{v \in V(G), i \in [k], n \in \mathbb{N}},$$

so that $X_v^{i,n}$ is at rate $n\sigma_v^i$. Then these countably many point processes can easily be seen to define the entire process; a spread event of type j from a vertex v to a vertex u that is currently in state ρ is controlled by the point process $X_{vu}^{j,\rho}$, with satisfaction events handled similarly.

An advantage of such a formulation is that it gives an easy coupling between processes that shows that if $B' \leq B$ pointwise, then the stochastic process $\text{DD}(\tau(0), B, \sigma, G)$ stochastically dominates $\text{DD}(\tau(0), B', \sigma, G)$ in the sense that in the coupling, the demands in the B process are always at least those in the B' process. This can be seen by noting that the rates satisfy $\rho B \geq \rho B'$ pointwise for all $\rho \in \mathbb{N}^l$. We thus take point processes $Y_e^{j,\rho}$ at rate $[\rho B - \rho B']_i$. If the point processes $\{X_e^{j,\rho}\}$ and $\{X_v^{i,n}\}$ are used to determine $\text{DD}(\tau(0), B, \sigma, G)$, then the point processes $\{X_e^{j,\rho} \cup Y_e^{j,\rho}\}$ and $\{X_v^{i,n}\}$ are used to determine $\text{DD}(\tau(0), B', \sigma, G)$.

In particular, this allows us to replace B with B' , where $B'_{ij} = \max\{B_{ij}, 0\}$, and conclusions about the extinction of the B' process still hold for B .

Furthermore, this turns out not to be entirely unreasonable; one hopes that the negative entries in B will afford better bounds on the extinction time, but in many cases with negative entries in B , extinctions of some demand types mean that the process is eventually run in a nonnegative case. In light of this, we will hereinafter assume for convenience that B is nonnegative.

Given an initial demand $\tau(0)$ and spread matrix B , our goal is to find a supply function σ such that demand is satisfied. Ideally, we would like to do this with small supply rates. Furthermore, the supply rates should depend only on the contact graph G , the spread matrix B , and the initial demand $\tau(0)$, but not on t or $\tau(t)$.

3. The Kronecker PageRank

The notion of PageRank was first introduced in [Brin and Page 98] for Google's search algorithms. Although PageRank was originally used for the Web graph, we can define the PageRank for any finite graph G . Here we will use a modified version of PageRank, called personalized PageRank, which has two parameters, a jumping constant $\alpha \in [0, 1]$ and a seed \mathbf{s} , which is some probability distribution on the vertex set V of G .

The personalized PageRank vector $\mathbf{pr}(\alpha, \mathbf{s})$ for jumping constant α and seed distribution \mathbf{s} on V is given by

$$\mathbf{pr}(\alpha, \mathbf{s}) = \alpha \sum_{l=0}^{\infty} (1 - \alpha)^l \mathbf{s} W^l.$$

Note that here we view \mathbf{s} as a row vector, which will be our convention for all vectors throughout this paper. We note that the PageRank vector is also the solution to the recurrence relation

$$\mathbf{pr}(\alpha, \mathbf{s}) = \alpha \mathbf{s} + (1 - \alpha) \mathbf{pr}(\alpha, \mathbf{s}) W.$$

The original definition of PageRank [Brin and Page 98] is the special case in which \mathbf{s} is the uniform distribution over all the vertices.

For a subset of vertices $H \subset V$, the volume of H is the sum of degrees of the vertices of H . The Cheeger ratio $h(H)$ of H measures the cut between H and its complement \bar{H} via the relationship

$$h(H) = \frac{e(H, \bar{H})}{\min\{\text{vol}(H), \text{vol}(\bar{H})\}}.$$

The α -core of a subset H is the set of vertices

$$C_\alpha = \left\{ v \in H \mid \mathbf{pr}(\alpha, 1_v)1_H \geq 1 - \frac{h}{\alpha} \right\}.$$

A basic tool for analyzing PageRank is the fact (see [Andersen et al. 07]) that for a subset with Cheeger ratio h , we can choose α , say $\alpha = h/2$, such that at least half of the vertices of H are in the α -core of H . Therefore, if the seed is in the α -core of H , then we can use PageRank to identify a large part of H . Another advantage of using PageRank is the fact that there are very efficient algorithms for approximating PageRank vectors [Andersen et al. 07].

One tool that we will use to understand the k -commodity dynamic demand model will be the Kronecker PageRank vector, which we define below. This is a generalization of the personalized PageRank vector.

Definition 3.1. (Kronecker PageRank.) Let B be a $k \times k$ matrix with spectral radius strictly less than 1, and let W be the transition matrix for a random walk on a graph G . Let \mathbf{s} be a nonnegative vector in $\mathbb{R}^{k \times |V|}$. The *Kronecker PageRank vector* with parameters B and \mathbf{s} is defined as

$$\mathbf{Kpr}(B, \mathbf{s}) = \sum_{l=0}^{\infty} \mathbf{s}(W \otimes B)^l = \sum_{l=0}^{\infty} \mathbf{s}(W^l \otimes B^l).$$

The requirement that the spectral radius of B be less than 1 is necessary to ensure convergence of the infinite sum, since the spectrum of $W \otimes B$ is the product of the spectra of W and B . Since the eigenvalues of W have absolute value at most 1, the sum will converge.

We note that when B is a 1×1 matrix $B = \beta < 1$ and \mathbf{s} is a probability distribution, then we have the relationship

$$\mathbf{Kpr}(B, \mathbf{s}) = \sum_{l=0}^{\infty} \mathbf{s}(W \otimes \beta)^l = \sum_{l=0}^{\infty} \mathbf{s}\beta^l W^l = \frac{1}{1-\beta} \mathbf{pr}(1-\beta, \mathbf{s}),$$

so the Kronecker PageRank is a natural extension of personalized PageRank. We will see in Theorem 5.3 that the Kronecker PageRank will arise naturally in our analysis in Section 3 and give better bounds than those afforded by standard PageRank by incorporating the spread matrix. We remark that the Kronecker PageRank vectors can be efficiently computed and approximated along the same lines as those for the usual PageRank.

For a square matrix A , there are many different matrix norms that can be used (see [Horn and Johnson 90]). We will use the following notation for the following norms:

1. $\|A\|_1 = \sum_{i,j} |a_{ij}|$ is the ℓ_1 norm.
2. $\|A\|_1 = \max_j \sum_i |a_{ij}|$ is the *maximum column sum norm*.
3. $\|A\|_\infty = \max_i \sum_j |a_{ij}|$ is the *maximum row sum norm*.
4. $\|A\|_2 = \max \left\{ \sqrt{\lambda} \mid \lambda \text{ is an eigenvalue of } A^*A \right\}$ is the *spectral norm*.

4. Global Analysis: Supplying Every Vertex

Here we show that if supply rates are above a certain threshold, then with probability approaching 1, demands will be satisfied.

Theorem 4.1. *Consider the k -commodity demand model on a graph G with n vertices parameterized by spread matrix $B = [\beta_{ij}]$. If the supply rates to each vertex v are*

$$\sigma_v^i > d_v \left(\sum_j \frac{\beta_{ij} + \beta_{ji}}{2} \right) + \delta$$

for $\delta > 0$, then with probability $1 - \epsilon$, all vertices are satisfied at time t for all

$$t > \frac{1}{\delta} \left(\frac{1}{2} \log(nk) + \log(X(0)) + \log \left(\frac{1}{\epsilon} \right) \right).$$

Proof. We consider the expectation $\mathbb{E}[\tau(t)]$. Let $X(t) = \|\tau(t)\|_1$, the total demand at time t . We will begin by considering the quantity $\frac{\partial}{\partial t} \mathbb{E}[\tau(t)]$.

From the discussion in Section 2, we know that demand is increasing with rates given by $\rho(t) = \tau(t)(A \otimes B)$, but also demand decreases according to the supply rates. Let $S = \text{diag}(\sigma)$, the diagonal $nk \times nk$ matrix with entries given by the supply vector. Then we can see that demand decreases at each vertex according to rates given by the supply rate vector $\tau(t)S$.

To proceed, we need the following two well-known and simple facts concerning exponentially distributed random variables. We use the notation $f(h) = 0_{h \rightarrow 0}(g(h))$ to indicate that $f(h) \leq Cg(h)$ for h sufficiently small.

Suppose X is an exponentially distributed waiting time with rate λ . Then

$$\mathbb{P}(X < h) = \lambda h + O_{h \rightarrow 0}(h^2),$$

which follows from the fact that the probability that an exponential waiting time is at most h is given by

$$\int_0^h \lambda e^{-\lambda h} = 1 - e^{-\lambda h} = \lambda h + O_{h \rightarrow 0}(h^2).$$

As an immediate consequence, we see that if X and Y are independent exponentially distributed waiting times with rates λ_1, λ_2 , then

$$\mathbb{P}(X, Y < h) = O_{h \rightarrow 0}(h^2). \quad (4.1)$$

Using these two facts, we will show that we can neatly encode the behavior of the k -commodity demand model in a single differential equation. Fix a vertex v and commodity i . We will show that

$$\frac{\partial}{\partial t} \mathbb{E}[\tau_v^i(t)] = [\mathbb{E}[\tau(t)](A \otimes B - S)]_v^i.$$

To do this, we compute the derivative by the definition; that is, we compute

$$\lim_{h \rightarrow 0} \frac{\mathbb{E}[\tau_v^i(t) - \tau_v^i(t+h)]}{h}.$$

To do so, consider the conditional expectation $\mathbb{E}[\tau_v^i(t) - \tau_v^i(t+h) \mid \tau(t)]$. Note that by (4.1), the probability that two independent events (two spread events, two satisfy events, or a spread and a satisfy event) occur is $O_{h \rightarrow 0}(h^2)$. On the other hand, given a neighbor u of v and a commodity j , the probability of a spread event originating from this neighbor and commodity in time $(t, t+h)$ is exactly $B_{ji}\tau_u^j(t)h + O_{h \rightarrow 0}(h^2)$. Likewise, the probability of a satisfaction event in this time is $\tau_v^i(t)\sigma_v^i h + O_{h \rightarrow 0}(h^2)$. Linearity of expectation yields

$$\mathbb{E}[\tau_v^i(t) - \tau_v^i(t+h) \mid \tau(t)] = \tau(t)(A \otimes B - S)h + o(h^2).$$

Applying the tower property of conditional expectation yields the result for this vertex v and commodity i . Since this holds for all choices of v and i , we obtain the single equation

$$\frac{\partial}{\partial t} \mathbb{E}[\tau(t)] = \mathbb{E}[\rho(t) - \tau(t)S] = \mathbb{E}[\tau(t)](A \otimes B - S).$$

Solving the matrix differential equation with initial condition $\mathbb{E}[\tau(0)] = \tau(0)$ yields

$$\mathbb{E}[\tau(t)] = \tau(0)e^{t(A \otimes B - S)}.$$

Let $Q = A \otimes B - S$. Then by [Dahlquist 59], $\|e^{tQ}\|_2 \leq e^{t\nu}$, where ν is the largest eigenvalue of $(Q + Q^*)/2$. We note that

$$\frac{Q + Q^*}{2} = A \otimes \left(\frac{B + B^*}{2} \right) - S,$$

which has diagonal terms $\beta_{ii} - \sigma_v^i$ ranging over all values of v and i . By the Gershgorin circle theorem, the eigenvalues of $(Q + Q^*)/2$ are contained in the

intervals

$$\left[-(d_v - 2)\beta_{ii} - d_v \left(\sum_{j \neq i} \frac{\beta_{ij} + \beta_{ji}}{2} \right) - \sigma_v^i, d_v \left(\sum_j \frac{\beta_{ij} + \beta_{ji}}{2} \right) - \sigma_v^i \right].$$

Since

$$\sigma_v^i > d_v \left(\sum_j \frac{\beta_{ij} + \beta_{ji}}{2} \right) + \delta,$$

all the eigenvalues of $(Q + Q^*)/2$ are less than $-\delta$. Therefore,

$$\begin{aligned} \mathbb{E}[X(t)] &= \left\| \tau(0) e^{t(A \otimes B - S)} \right\|_1 \leq \sqrt{nk} \left\| \tau(0) e^{t(A \otimes B - S)} \right\|_2 \\ &\leq \sqrt{nk} \|\tau(0)\|_2 \left\| e^{t(A \otimes B - S)} \right\|_2 \leq \sqrt{nk} \|\tau(0)\|_1 e^{t\nu} \leq \sqrt{nk} X(0) e^{-t\delta}. \end{aligned}$$

Thus Markov's inequality gives that $\mathbb{P}(X(t) > 0) < \epsilon$ if

$$t > \frac{1}{\delta} \left(\frac{1}{2} \log(nk) + \log(X(0)) + \log\left(\frac{1}{\epsilon}\right) \right).$$

□

We note that this approach works for all initial distributions $\tau(0)$. This indicates that in many situations, this approach may be overkill and that we could have used smaller supply rates. In the next section, we analyze the process more carefully and give conditions that depend on the initial distribution of demand.

5. Local Analysis: Supplying a Small Subset

For the remainder of the discussion, it is convenient to introduce a reformulation of the model that takes advantage of the fact that demands take on integer values. Rather than view demands as a function $\tau : V \rightarrow \mathbb{N}^k$, we view demands as discrete objects sitting on each node. Borrowing language from chip-firing games on graphs (see, for example, [Merino 05]), we view units of the demand as chips located on vertices of the graph. For example, if $k = 7$ and for a vertex v , we have $\tau_v(t) = (0, 1, 2, 0, 2, 0, 3)$, then we would say that at time t , there were one 2-chip, two 3-chips, two 5-chips, and three 7-chips at vertex v , corresponding to one “unit of demand” for commodity 2, and so on.

In contrast to classical chip-firing games, the number of chips is not static, and the game is parameterized by continuous time. We restate the possible transitions in terms of demand chips. For an i -chip at vertex v , there are two types of transition events:

1. For each vertex $w \sim v$ and each $j = 1, \dots, l$, a j -chip is added at w at rate β_{ij} . When this occurs, we say that the new j -chip is created by the i -chip.
2. The i -chip itself is removed at rate σ_v^i .

It is important to note that due to the properties of exponential random variables, the rates add linearly, and the model is equivalent to the original description discussed in Section 2. The main advantage of this reformulation is the ability to trace back the history of a chip. If there is a chip c_l at vertex v_l at time t , then either c existed at time $t = 0$ or there is a sequence of $l + 1$ chips (c_0, \dots, c_l) located at vertices along a path $\pi = (v_0, v_1, \dots, v_l)$, where c_0 existed at $t = 0$, and c_r is created by c_{r-1} for $r = 1, \dots, l$. We allow π to have repeated vertices to allow for the case in which demand created more demand at the same vertex. If a chip c exists at time 0, we refer to it as an *initial chip*.

For a path $\pi = (v_0, v_1, \dots, v_l)$ and a chip c_0 located at v_0 , we define the event \mathcal{S}_{π, c_0} to be the event that there is a sequence of $l + 1$ chips (c_0, \dots, c_l) located respectively at (v_0, v_1, \dots, v_l) , and c_r is created by c_{r-1} for $r = 1, \dots, l$.

It is important to note that \mathcal{S}_{π, c_0} occurring does not imply that there is any demand at v_l at time t , because it could have been satisfied at some time before t . However, if there is a demand at v_l at time t , then $\mathcal{S}_{\pi, c}$ must have occurred for some initial chip c at vertex v_0 and some walk π from v_0 to v_l .

We begin by showing that if supply rates are large enough, then the probability of demands spreading along a long walk is small. Inspired by Theorem 4.1, we make the assumption that supply rates are proportional to the degree of the vertices. That is, we assume that $\sigma_v^i > \mu_i(d_v)$ for all v for constants $\mu_i > 0$.

Lemma 5.1. *Let $M = \text{diag}(\mu_1, \dots, \mu_k)$, $\hat{B} = M^{-1}B$ and $\zeta = \min\{\|\hat{B}\|_1, \|\hat{B}\|_\infty\}$. Then for every chip c_0 located at v_0 and every walk $\pi = (v_0, \dots, v_l)$ of length l , we have*

$$P(\mathcal{S}_{\pi, c_0}) \leq k \prod_{j=0}^l \frac{1}{d_{v_j}} \zeta^l.$$

Proof. Let S_r denote the event that a chip c_r at v_r creates a chip at v_{r+1} . If c_r is an i -chip, then for it to create any chip at v_{r+1} , a spread event must occur before c_r is removed. The time until c_r creates a j -chip at v_{r+1} is an exponential random variable with rate β_{ij} . Since the time until c_r is removed is given by

$\text{Exp}(\sigma_v^i)$, the probability of c_r creating a j -chip is

$$\frac{\beta_{ij}}{\beta_{ij} + \sigma_v^i} \leq \frac{\beta_{ij}}{\sigma_v^i} < \frac{\beta_{ij}}{\mu_i d_{v_r}}.$$

Thus

$$\mathbb{P}(S_r) < \sum_{i,j} \frac{\beta_{ij}}{\mu_i d_{v_r}} = \frac{1}{d_{v_r}} \mathbf{1} \hat{B} \mathbf{1}^*.$$

For a walk π of length l , we want to consider the intermediate steps more carefully. Since there are l transitions that occur, we can use the same reasoning as above to obtain the bound

$$\mathbb{P}(S_{\pi,c}) < \prod_{r=0}^l \frac{1}{d_{v_r}} \mathbf{1} \hat{B}^l \mathbf{1}^* = \prod_{r=0}^l \frac{1}{d_{v_r}} \|\hat{B}^l\|_1 \leq k \prod_{r=0}^l \frac{1}{d_{v_r}} \|\hat{B}^l\|_1 \leq k \prod_{r=0}^l \frac{1}{d_{v_r}} \|\hat{B}\|_1^l.$$

The factor of k that appears in the final lines above is just a consequence of switching from the vector 1-norm $\|\hat{B}^l\|_1$ to the maximum column sum norm $\|\hat{B}^l\|_1$ (see [Horn and Johnson 90]).

We could just as easily have switched to the maximum row sum norm and obtained the term $k \|\hat{B}\|_\infty^l$, and so it follows that

$$\mathbb{P}(S_{\pi,c}) < \min \left\{ k \prod_{r=0}^l \frac{1}{d_{v_r}} \|\hat{B}\|_1^l, k \prod_{r=0}^l \frac{1}{d_{v_r}} \|\hat{B}\|_\infty^l \right\} = k \prod_{j=0}^l \frac{1}{d_{v_j}} \zeta^l.$$

□

We note that the decision to use $\zeta = \min\{\|\hat{B}\|_1, \|\hat{B}\|_\infty\}$ in Lemma 5.1 reflects the difficulty in working with arbitrary spread matrices B . For certain classes of spread matrices (e.g., if B is symmetric or diagonalizable), it is possible to obtain tighter bounds. While the previous lemma will be useful in obtaining a bound using PageRank, a more careful analysis is possible that will lead naturally to use of Kronecker PageRank, which we explore in Theorem 5.3.

Theorem 5.2. *Suppose that initial demand is contained in $S \subset H \subset V$, and each vertex $v \in H$ has supply rates $\sigma_v^i > \mu_i d_v$ and $\sigma_w^i = 0$ for $w \in \bar{H}$. Let $M = \text{diag}(\mu_1, \dots, \mu_k)$, $\hat{B} = M^{-1}B$, and $\zeta = \min\{\|\hat{B}\|_1, \|\hat{B}\|_\infty\}$. Let $\mathbf{x}(t)$ be defined by $x_v(t) = \sum_i \tau_v^i(t)$, and $X(t) = \|\mathbf{x}(t)\|_1 = \|\boldsymbol{\tau}(t)\|_1$. Let E_H denote the event that demands spread outside the set H . Then:*

1. $\mathbb{P}(E_H) \leq \frac{X(0)}{\zeta} \mathbf{pr} \left(1 - \zeta, \frac{\mathbf{x}(0)}{X(0)} \right) \mathbf{1}_H^*.$
2. *If S is in the $(1 - \zeta)$ core of H , then $\mathbb{P}(E_H) \leq \frac{2X(0)h(H)}{\zeta(1-\zeta)}$, where $h(H)$ is the Cheeger ratio of H .*

Proof. Let P_l denote the set of all paths of length l from an initial chip in S to \bar{H} such that the first $k-1$ steps are in H . Let $P = \bigcup_{l=1}^{\infty} P_l$. The key observation is that if $w \in \bar{H}$ ever has demand, then $S_{\pi,c}$ must have occurred for some initial chip c and path π from the location of c to w . Thus we can use the union bound to get that

$$\begin{aligned}
\sum_{\pi \in P} \mathbb{P}(S_{\pi,c}) &\leq \sum_l \sum_{(\pi,c) \in P_l} \mathbb{P}(S_{\pi,c}) \\
&\leq \sum_l \sum_{v_0 \in S} \sum_{c \text{ at } v_0} \sum_{v_l \in \bar{H}} \sum_{\pi=(v_0, \dots, v_l) \in P_l} \mathbb{P}(S_{\pi,c}) \\
&\leq \sum_l \sum_{v_0 \in S} \sum_{c \text{ at } v_0} \sum_{v_l \in \bar{H}} \sum_{\pi=(v_0, \dots, v_l) \in P_l} \zeta^l \prod_{r=0}^l \frac{1}{d_{v_r}} \\
&= \sum_l \mathbf{x}(0) \zeta^l (D^{-1}A)^l \mathbf{1}_{\bar{H}}^* = \sum_l \mathbf{x}(0) \zeta^l W^l \mathbf{1}_{\bar{H}}^* \\
&= \frac{X(0)}{\zeta} \mathbf{pr} \left(1 - \zeta, \frac{\mathbf{x}}{X(0)} \right) \mathbf{1}_{\bar{H}}^*,
\end{aligned}$$

proving the first statement. The second statement follows the same proof as [Chung et al. 09, Theorem 3.2]. \square

Theorem 5.3. *Suppose that the initial demand is contained in $S \subset H \subset V$, and each vertex $v \in H$ has supply rates $\sigma_v^i \geq \mu_i d_v$. Let $M = \text{diag}(\mu_1, \dots, \mu_k)$, $\hat{B} = M^{-1}B$, and $\zeta = \|\hat{B}\|_1$. Let $X(t) = \|\tau(t)\|_1$, the total number of demands at time t . Let \mathcal{E}_H denote the event that demands spread outside the set H . Then \mathcal{E}_H can be bounded above using the Kronecker PageRank vector via the relationship*

$$\mathbb{P}(\mathcal{E}_H) \leq X(0) \mathbf{Kpr} \left(\hat{B}, \frac{\tau(0)}{X(0)} \right) \mathbf{1}_{\bar{H}}.$$

Proof. Let f be a vector indicator function of commodity type on chips, that is, $f(c) = \mathbf{e}_i$ if c is an i -chip, where \mathbf{e}_i denotes the i th standard basis vector for \mathbb{R}^k . Let C_0 denote the set of initial chips. By the same methods that were used in the proof of Lemma 5.1, we can bound the probability that demand originating from c ever spreads along a path $\pi = (v_0, v_1, \dots, v_l)$ by the sum

$$\mathbb{P}(S_{\pi,c}) \leq f(c) \hat{B}^l \mathbf{1}^* \prod_{r=0}^l \frac{1}{d_{v_r}}.$$

Therefore, using the same technique as in the proof of Theorem 5.2, we obtain the bound

$$\begin{aligned}
\sum_{\pi \in P} \mathbb{P}(\mathcal{S}_{\pi,c}) &\leq \sum_l \sum_{u \in S} \sum_{\pi \in B_l} \mathbb{P}(\mathcal{S}_{\pi,u}) \leq \sum_l \sum_{c \in C_0} \sum_{v_l \in \bar{H}} \sum_{\pi=(v_0, \dots, v_l) \in P_l} \mathbb{P}(\mathcal{S}_{\pi,u}) \\
&\leq \sum_l \sum_{c \in C_0} \sum_{v_l \in \bar{H}} \sum_{\pi=(v_0, \dots, v_l) \in P_l} f(c) \hat{B}^l \mathbf{1}^* \prod_{r=0}^l \frac{1}{d_{v_r}} \\
&= \sum_l \tau(0) (D^{-1} A \otimes \hat{B})^l \mathbf{1}_{\bar{H}} \\
&= \sum_l \tau(0) (W \otimes \hat{B})^l \mathbf{1}_{\bar{H}} = X(0) \mathbf{Kpr} \left(\hat{B}, \frac{\tau(0)}{X(0)} \right) \mathbf{1}_{\bar{H}}.
\end{aligned}$$

□

Let \mathcal{S}_t denote the event that all of the vertices are satisfied at time t . In order to complete the analysis of the local case, we would like to bound $\mathbb{P}(\mathcal{S}_t | \bar{\mathcal{E}}_H)$, where \mathcal{E}_H is as in Theorems 5.2 and 5.3. Such a bound is not immediately given by Theorem 4.1, but obtaining one is not difficult. To derive a bound on $\mathbb{P}(\mathcal{S}_t | \bar{\mathcal{E}}_H)$, consider running a modified Dirichlet version that is identical to the standard process with the same supply rates, except demand leaving H is ignored. Let \mathcal{S}'_t denote the event that in the Dirichlet process, all of the events are satisfied at time t . Then $\mathbb{P}(\mathcal{S}'_t)$ can be bounded directly by Theorem 4.1, since this Dirichlet process restricted to vertices in H is the standard process on H . Furthermore, $\mathbb{P}(\mathcal{S}_t \cap \mathcal{E}_H) \leq \mathbb{P}(\mathcal{S}'_t)$. Therefore,

$$\mathbb{P}(\mathcal{S}_t | \bar{\mathcal{E}}_H) = \frac{\mathbb{P}(\mathcal{S}_t \cap \bar{\mathcal{E}}_H)}{\mathbb{P}(\bar{\mathcal{E}}_H)} \leq \frac{\mathbb{P}(\mathcal{S}'_t)}{\mathbb{P}(\bar{\mathcal{E}}_H)}.$$

Combining this observation along with Theorems 5.2 and 5.3 yields that the probability of escape from H is bounded, and if the process does not escape from H , it dies quickly.

Theorems 5.2 and 5.3 can be used in two different ways. As stated, they provide a way to bound the probability that demands escape from a given subset. However, they can be also used to construct such a bounding subset. For example, given initial demand $\tau(0)$ contained in an initial set of vertices $S \subset V$, we can algorithmically construct H such that demand stays in H with probability $1 - \epsilon$ as follows. We do this by constructing an increasing family of subsets $\{H_r\}$. We begin by setting $H_0 = S$ and

$$\mathbf{pr} = \frac{X(0)}{\zeta} \mathbf{pr} \left(1 - \zeta, \frac{\tau(0)}{X(0)} \right).$$

Then we follow the following procedure:

1. Compute $f(r) = \langle \mathbf{pr}, \mathbf{1}_{\bar{H}_r}^* \rangle$.
2. If $f(r) < \epsilon$, set $H = H_r$ and end the process.
3. If $f(r) \geq \epsilon$, let $H_{r+1} = H_r \cup v$ for some $v \notin H_r$, and return to step 1.

This process will eventually terminate, since $|H_{r+1}| = |H_r| + 1$, and $f(r) = 0$ once $H_r = V$.

6. An Example on a Random Geometric Graph

We conclude with an example calculation on synthetic data. Our graph G is an instance of a random geometric graph.

Two hundred vertices were placed uniformly at random in a unit square, and two vertices are adjacent if the distance between them is less than 0.13 units.

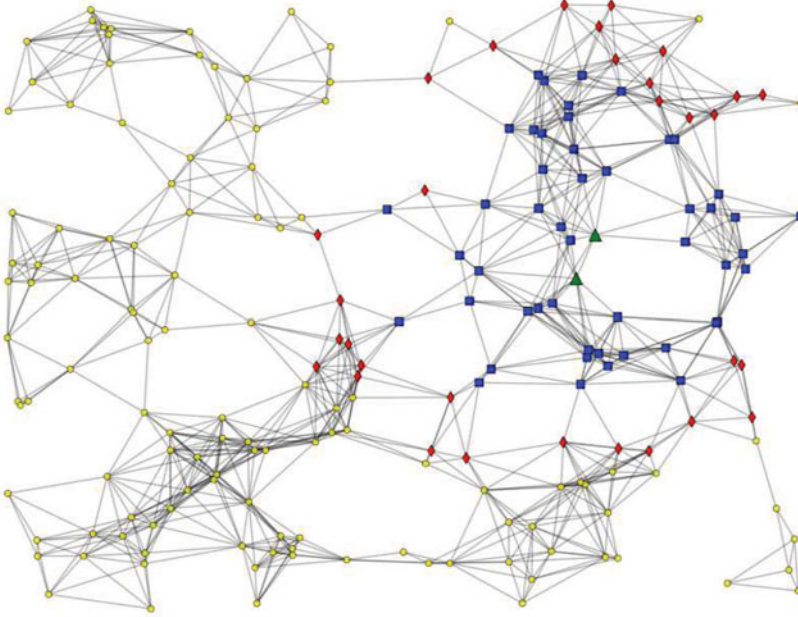


Figure 1. A comparison between Theorems 5.2 and 5.3 on a random geometric graph.

We let $k = 3$ and

$$B = \begin{pmatrix} 0.4 & 0.2 & 0.3 \\ 0.2 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.5 \end{pmatrix}.$$

The initial demand is given by $\tau_v^j(0) = 1$ for all commodities for the vertices marked by triangles, and $\tau_v^j(0) = 0$ for all other vertices. In addition, we set $\mu_i = 3$ for all i .

We demonstrate the difference between Theorems 5.2 and 5.3 in the following way. Figure 1 shows the graph G . The demands start in the triangular vertices and spread outward from there. Theorem 5.3 states that with 95% probability, demands stay in the square vertices. Theorem 5.2 states that with 95% probability, demands stay in the diamond and square vertices. This small example illustrates how Kronecker PageRank can be used to obtain improved results.

References

- [Andersen et al. 07] R. Andersen, F. Chung, and K. Lang. “Local Partitioning for Directed Graphs Using PageRank.” In *Proceedings of the 5th International Conference on Algorithms and Models for the Web-Graph*, pp. 166–178, 2007.
- [Borgs et al. 10] Christian Borgs, Jennifer Chayes, Ayalvadi Ganesh, and Amin Saberi. “How to Distribute Antidote to Control Epidemics.” *Random Structures and Algorithms* 37 (2010), 204–222.
- [Brin and Page 98] S. Brin and L. Page. “The Anatomy of a Large-Scale Hypertextual Web Search Engine.” *Computer Networks and ISDN Systems* 30 (1998), 107–117.
- [Chevalleyre et al. 06] Y. Chevalleyre, P. E. Dunne, U. Endriss, J. Lang, M. Lemaitre, N. Maudet, J. Padget, S. Phelps, J. A. Rodriguez-Aguilar, and P. Sousa. “Issues in Multiagent Resource Allocation.” *Informatica* 30 (2006), 3–31.
- [Chung et al. 09] Fan Chung, Paul Horn, and Alexander Tsiatas. “Distributing Antidote Using PageRank Vectors.” *Internet Mathematics* 6 (2009), 237–254.
- [Dahlquist 59] Germund Dahlquist. “Stability and Error Bounds in the Numerical Integration of Ordinary Differential Equations.” *Kungl. Tekn. Högsk. Handl. Stockholm*. 130 (1959), 267–268.
- [Ganesh et al. 05] A. Ganesh, L. Massoulie, and D. Towsley. “The Effect of Network Topology on the Spread of Epidemics.” *IEEE* 2 (2005), 1455–1466.
- [Horn and Johnson 90] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

- [Kiss et al. 06] Istvan Z Kiss, Darren M Green, and Rowland R Kao. “Infectious Disease Control Using Contact Tracing in Random and Scale-Free Networks.” *Journal of the Royal Society* 3 (2006), 55–62.
- [Merino 05] C Merino. “The Chip-Firing Game.” *Discrete Mathematics* 302 (2005), 188–210.
- [Newman 02] M. Newman. “Spread of Epidemic Disease on Networks.” *Physical Review E* 66 (2002), 016128.

Fan Chung, Department of Mathematics, University of California, San Diego, 9500 Gilman Drive # 0112, La Jolla, CA 92093-0112, USA (fan@math.ucsd.edu)

Paul Horn, Department of Mathematics, Harvard University, Cambridge, Massachusetts, USA. Current address: Department of Mathematics, University of Denver, Aspen Hall, Room 717, 2280 S. Vine Street, Denver, CO 80208, USA (paul.horn@du.edu)

Jacob Hughes, Department of Mathematics, University of California, San Diego, 9500 Gilman Drive # 0112, La Jolla, CA 92093-0112, USA (jthughes@math.ucsd.edu)