Doron Shahar

RTG Report: Modeling web-crawlers on the Internet with Random walks on graphs

This report focuses on the results developed in the article "On certain connectivity properties of the Internet topology". I have written an introduction to the topics necessary for understanding the paper, provided a more through explanation of topics and proofs, improved on the key result presented in the paper, and have proven an opposing result about one case not covered in the theorem.

To begin, I will note one notation I shall be using. For a matrix A, $A(i,j)$ will denote the (i,j)th entry of A. Similarly, $[AB](i,j)$ will denote the (i,j)th entry of AB. A similar notation will be used for vectors. If v is a vector, $v(i)$ will be the ith coordinate of v. And $[vA](i)$ will be the ith coordinate of the vector obtained by multiplying v by A on the left.

Note that we will use the word graph in a less restrictive sense to include multiigraphs. That is, our graphs may have multiple edges connecting two vertices. Furthermore, self-loops will contribute two to the degree of a vertex.

1 Motivation

The state of the Internet at any time can be modeled by a graph where websites are the vertices and the links between websites are the edges of the graph. As websites are formed and new links are made the graph representing the Internet changes. As such, we can model the way in which the Internet changes over time by a graph that changes over time.

Some simple models for how the Internet grow are to assume that each new website links to some of the websites already present with a given probability and that no other links are formed. These are the models used in the paper "On certain connectivity properties of the Internet topology". Our graphs then that model the Internet are random graphs.

Our key interest shall be in web crawlers. A web crawler browses the Internet for the purpose of indexing websites. This is one the way in which search engines find new sites to list. We can model the way in which the web crawler browses by a random walk on a graph.

My goal then is to describe properties of random walks on the random graphs that model the growth of the Internet.

2 Random walks and Reversible Markov chains

We shall be dealing with random walks on connected graphs for which there is a unique probability distribution π. Throughout the paper, we will be considering a random walk on a connected graph with n vertices with transition matrix P and a unique probability distribution π. As we shall show shortly, every random walk on a graph is a reversible Markov chain. Many of the key results are proven for reversible Markov chains.

Definition: A Markov Chain with transition matrix P is reversible if there is probability distribution π such that $\pi(i)P(i,j) = \pi(j)P(j,i)$.

Proposition 1: *For a reversible Markov chain with transition matrix P, the probability distribution π such that $\pi(i)P(i,j) = \pi(j)P(j,i)$ is an invariant probability distribution of the Markov chain.*

Proof: $[\pi P](j) = \sum_i \pi(i)P(i,j) = \sum_i \pi(j)P(j,i) = \pi(j)\sum_i P(j,i) = \pi(j)$. Therefore, $\pi P = \pi$. So π is an invariant probability distribution of the Markov chain. QED

Proposition 2: *Every random walk on a graph is a reversible Markov chain.*

Proof: Let $\deg(i)$ be the degree of the ith vertex, and let $e(i,j)$ be the number of edges connecting the ith and jth vertex. $\sum_j e(i,j) = \deg(i)$ and $e(i,j) = e(j,i)$. $\pi(i) = \frac{\deg(i)}{\sum_k \deg(k)}$. $P(i,j) = \frac{e(i,j)}{\deg(i)}$. Therefore, $\pi(i)P(i,j) = \frac{\deg(i)}{\sum_k \deg(k)}\frac{e(i,j)}{\deg(i)} = \frac{e(i,j)}{\sum_k \deg(k)} = \frac{e(j,i)}{\sum_k \deg(k)} = \frac{\deg(j)}{\sum_k \deg(k)}\frac{e(j,i)}{\deg(j)} = \pi(j)P(j,i)$

So every random walk on a graph is a reversible Markov chain. QED

The following matrices will be useful when proving facts about reversible Markov chains. Let D be the diagonal matrix with $D(i,i) = \pi(i)$, $Q = DP$, and $A = D^{1/2}PD^{-1/2}$.

Q and A are symmetric matrices.

Proposition 3: *Q is a symmetric matrix.*

Proof: $Q(i,j) = [DP](i,j) = \sum_k D(i,k)P(k,j) = D(i,i)P(i,j) = \pi(i)P(i,j) = \pi(j)P(j,i) = D(j,j)P(j,i) = \sum_k D(j,k)P(k,i) = [DP](j,i) = Q(j,i)$. QED

Proposition 4: *A is a symmetric matrix*

Proof: $A(i,j) = \sqrt{\pi(i)}P(i,j)\frac{1}{\sqrt{\pi(j)}} = \frac{1}{\sqrt{\pi(i)}}\pi(i)P(i,j)\frac{1}{\sqrt{\pi(j)}} = \frac{1}{\sqrt{\pi(i)}}\pi(j)P(j,i)\frac{1}{\sqrt{\pi(j)}} = \sqrt{\pi(j)}P(j,i)\frac{1}{\sqrt{\pi(i)}} = A(j,i)$. QED

3 Convergence to stationary state

In the model presented the paper, the random walk on the graph is irreducible because the graph is connected, and it is aperiodic because the first vertex has a selfloop. So by the ergodic theorem, P^t ($t \in \mathbb{N}$) converges to the matrix Π whose rows are all π. Physically, this means that for large time t the initial state of the system does not influence the final state of the system. In regards to the Internet, this means that eventually the probability that the web crawler starts visits a particular website will be the same regardless of which website it started at. This "eventually" is a time t when $P^t \approx \Pi$. It is related to the time when it is likely that the web-crawler has indexed every website.

This is our motivation for wanting to know how quickly P^t converges to Π. The faster the convergence the sooner the web-crawler will index all the websites. We will measure convergence using two common measures: the relative pointwise distance and the total variation distance. Note that the relative pointwise distance is larger than the total variation distance. Hence, convergence in the former guarantees convergence in the latter.

Definition:

1. The relative pointwise distance is $\Delta(t) = \max_{i,j} |\frac{P^t(i,j)}{\pi(j)} - 1| = \max_{i,j} |\frac{P^t(i,j)}{\Pi(i,j)} - 1|$.

2. The total variation distance is $\delta(t) = \max_i \sum_j |P^t(i,j) - \pi(j)| = \max_i \sum_j |P^t(i,j) - \Pi(i,j)|$.

Next, let's verify the claim that the relative pointwise distance is indeed larger than the total variation distance.

Proposition 5: $\Delta(t) \geq \delta(t)$

Proof: $\Delta(t) = \max_{i,j} |\frac{P^t(i,j)}{\pi(j)} - 1| \geq |\frac{P^t(i,j)}{\pi(j)} - 1|$ for all i, j. Therefore, $\Delta(t) \geq \sum_j |\frac{P^t(i,j)}{\pi(j)} - 1|\pi(j)$ for all i, because the right hand side is a weighted average of quantities that are all less than or equal to $\Delta(t)$. So $\Delta(t) \geq \max_i \sum_j |\frac{P^t(i,j)}{\pi(j)} - 1|\pi(j) = \max_i \sum_j |P^t(i,j) - \pi(j)| = \delta(t)$. QED

The following lemma gives a simple formula for P^t that is used to find how quickly it converges to Π. The idea of the proof it to use reversibility to relate P to a symmetric matrix A, the matrix A defined earlier. Then the spectral theorem can be used to write A in a convenient form. We do not want to diagonalize A, because it is unclear what the matrix, which diagonalizes A, will look like. So it is unclear what A^t will be if we first diagonalize A. Once A is written in a nice form, it is simple to calculate its powers and to relate the result back to P.

Note that the eigenvalues of P are $\{\lambda_1, \lambda_2, ..., \lambda_n\}$ where $1 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ..., \lambda_n \geq -1$. This should follow from a proof similar to that of the Perron-Frobenius Theorem.

For the proof we shall also need the matrix D described earlier. Recall that D is diagonal matrix whose diagonal entries are $\pi(i)$.

Lemma 6: $P^t(i,j) = \pi(j) + \sqrt{\frac{\pi(j)}{\pi(i)}} \sum_{m=2}^{n} \lambda_m^t e_m(i) e_m(j)$ *where* $\{e_1, ..., e_n\}$ *are orthonormal vectors and* $\{\lambda_1, \lambda_2, ..., \lambda_n\}$ *are the eigenvalues of* P.

Proof: The matrix $A = D^{1/2} P D^{-1/2}$ was shown to be symmetric for a reversible Markov Chain. Note that A and P have the same eigenvalues, because they are similar matrices. Since A is symmetric, it has an orthonormal basis of eigenvectors $\{e_1, ..., e_n\}$. The matrix A can be viewed as a linear transformation. Hence, it is defined by how it acts on the e_i: $Ae_i = \lambda_i e_i$. Using the fact that the e_i are orthonormal we can find another representation of A. Consider the $n \times n$ matrices $\lambda_m e_m e_m^T$ for $m = 1, ..., n$. These matrices are chosen because multiplying e_l by them results in an inner product: $(\lambda_m e_m e_m^T) e_l = \lambda_m e_m \langle e_m, e_l \rangle = \lambda_m e_m \delta_{lm} = \lambda_m e_m$ if $m = l$ and is zero otherwise. So $\lambda_m e_m e_m^T$ has eigenvector e_m with eigenvalue λ_m. All other eigenvectors of A are in the kernel of $\lambda_m e_m e_m^T$. Therefore, $\sum_{m=1}^{n} \lambda_m e_m e_m^T$ has eigenvectors $e_1, ..., e_m$ whose corresponding eigenvalues are $\lambda_1, ..., \lambda_n$. So $\sum_{m=1}^{n} \lambda_m e_m e_m^T$ acts on the basis $\{e_1, ..., e_m\}$ exactly as A does. That is, $A = \sum_{m=1}^{n} \lambda_m e_m e_m^T$. This is the convenient form in which we want A.

We can now calculate the powers of A. It is more instructive, I think, to first see how A^2 is calculated. D^t can then be calculated by induction, or directly using an argument like the one shown for A^2 but with more indices.

$$A^2 = (\sum_{m=1}^{n} \lambda_m e_m e_m^T)(\sum_{l=1}^{n} \lambda_l e_l e_l^T) = \sum_{m=1}^{n}\sum_{l=1}^{n} \lambda_m \lambda_l e_m e_m^T e_l e_l^T = \sum_{m=1}^{n}\sum_{l=1}^{n} \lambda_m \lambda_l e_m \langle e_m, e_l \rangle e_l^T$$
$$= \sum_{m=1}^{n}\sum_{l=1}^{n} \lambda_m \lambda_l e_m \delta_{m,l} e_l^T = \sum_{m=1}^{n} \lambda_m \lambda_m e_m e_m^T = \sum_{m=1}^{n} \lambda_m^2 e_m e_m^T$$

The key aspect of the calculation is the the cross terms in the multiplication cancel because of the orthogonality of the vectors. The same holds when taking higher powers of A. In general, $A^t = \sum_{m=1}^{n} \lambda_m^t e_m e_m^T$.

Finally, we can relate the result back to P.
$P^t = (D^{-1/2}AD^{1/2})^t = D^{-1/2}A^tD^{1/2} = D^{-1/2}(\sum_{m=1}^{n} \lambda_m^t e_m e_m^T)D^{1/2} = \sum_{m=1}^{n} \lambda_m^t D^{-1/2} e_m e_m^T D^{1/2}$.

The (i,j)th entry of $D^{-1/2}e_m e_m^T D^{1/2}$ is ith coordinate of the vector $D^{-1/2}e_m$ times the jth coordinate of the vector $e_m^T D^{1/2}$. The ith coordinate of the vector $D^{-1/2}e_m$ is $\frac{1}{\sqrt{\pi(i)}}e_m(i)$, and the jth coordinate of the vector $e_m^T D^{1/2}$ is $\sqrt{\pi(j)}e_m(j)$. Therefore, the (i,j)th coordinate of $D^{-1/2}e_m e_m^T D^{1/2}$ is $\sqrt{\frac{\pi(j)}{\pi(i)}}e_m(i)e_m(j)$.

$P^t(i,j) = \sum_{m=1}^{n} \lambda_m^t (D^{-1/2}e_m e_m^T D^{1/2})(i,j) = \sum_{m=1}^{n} \lambda_m^t \sqrt{\frac{\pi(j)}{\pi(i)}} e_m(i)e_m(j) = \sqrt{\frac{\pi(j)}{\pi(i)}} \sum_{m=1}^{n} \lambda_m^t e_m(i)e_m(j)$.

It would be helpful when evaluating $\Delta(t)$ or $\delta(t)$, if the above expression had the term $\pi(j)$ in it. Nice cancellations will occur if this is the case. This will indeed be possible by relating e_1 (an eigenvector of A with eigenvalue $\lambda_1 = 1$) with π (an eigenvector of P with eigenvalue $\lambda_1 = 1$).

$\pi^T = (\pi P)^T = P^T \pi^T = (D^{-1/2}AD^{1/2})^T \pi^T = D^{1/2}AD^{-1/2}\pi^T$, because A is symmetric. So $D^{-1/2}\pi^T = AD^{-1/2}\pi^T$. Therefore, $D^{-1/2}\pi^T$ is an eigenvector of A with eigenvalue 1. Furthermore, $D^{-1/2}\pi^T = (\sqrt{\pi(1)}, ..., \sqrt{\pi(n)})^T$ has norm 1. Note that both A and P may have multiple eigenvectors with eigenvalue 1 because it may be the case that $1 = \lambda_1 = \lambda_2 = \cdots$. But by reordering the basis $\{e_1, ..., e_m\}$, we may choose to have $e_1 = D^{-1/2}\pi^T$. Thus, $e_1(i) = (D^{-1/2}\pi)(i) = \sqrt{\pi(i)}$.

Therefore,

$$\begin{aligned}
P^t(i,j) &= \sqrt{\frac{\pi(j)}{\pi(i)}} \sum_{m=1}^{n} \lambda_m^t e_m(i)e_m(j) \\
&= \sqrt{\frac{\pi(j)}{\pi(i)}} (\lambda_1^t e_1(i)e_1(j) + \sum_{m=2}^{n} \lambda_m^t e_m(i)e_m(j)) \\
&= \sqrt{\frac{\pi(j)}{\pi(i)}} (\sqrt{\pi(i)}\sqrt{\pi(j)} + \sum_{m=2}^{n} \lambda_m^t e_m(i)e_m(j)) \\
&= \pi(j) + \sqrt{\frac{\pi(j)}{\pi(i)}} \sum_{m=2}^{n} \lambda_m^t e_m(i)e_m(j))
\end{aligned}$$

QED

We can now bound $\Delta(t)$ and $\delta(t)$. The first bound can be found in Durrett. I tried to find a second better bound for $\delta(t)$, but the easiest bound I could find was the same one that bounds $\Delta(t)$. Such a bound follows, because $\Delta(t) \geq \delta(t)$. The problem is that the sum in the definition of $\delta(t)$ cannot be easily bounded without needing to take a maximum.

We shall need the following definition for the proof.

Definition: $\lambda_{max} = \max\{|\lambda_2|, |\lambda_n|\} = \max\{|\lambda_2|, |\lambda_3|, \ldots, |\lambda_n|\}$

Proposition 7: $\Delta(t) \leq \frac{1}{\min_k \pi(k)} \lambda_{max}^t$

Proof:

$$
\begin{aligned}
\Delta(t) &= \max_{i,j} \Big| \frac{P^t(i,j)}{\pi(j)} - 1 \Big| \\
&= \max_{i,j} \Big| \frac{\pi(j) + \sqrt{\frac{\pi(j)}{\pi(i)}} \sum_{m=2}^n \lambda_m^t e_m(i) e_m(j)}{\pi(j)} - 1 \Big| && \text{by the lemma} \\
&= \max_{i,j} \Big| \frac{\sqrt{\frac{\pi(j)}{\pi(i)}} \sum_{m=2}^n \lambda_m^t e_m(i) e_m(j)}{\pi(j)} \Big| \\
&= \max_{i,j} \Big| \frac{\sum_{m=2}^n \lambda_m^t e_m(i) e_m(j)}{\sqrt{\pi(i)\pi(j)}} \Big| \\
&\leq \max_{i,j} \frac{\sum_{m=2}^n |\lambda_m|^t |e_m(i)||e_m(j)|}{\sqrt{\pi(i)\pi(j)}} \\
&\leq \frac{\max_{i,j} \sum_{m=2}^n |\lambda_m|^t |e_m(i)||e_m(j)|}{\min_{i,j} \sqrt{\pi(i)\pi(j)}} \\
&\leq \frac{\max_{i,j} \sum_{m=2}^n \lambda_{max}^t |e_m(i)||e_m(j)|}{\min_k \sqrt{\pi(k)\pi(k)}} \\
&= \frac{1}{\min_k \pi(k)} \lambda_{max}^t \max_{i,j} \sum_{m=2}^n |e_m(i)||e_m(j)| \\
&\leq \frac{1}{\min_k \pi(k)} \lambda_{max}^t \max_{i,j} \sum_{m=1}^n |e_m(i)||e_m(j)| \\
&\leq \frac{1}{\min_k \pi(k)} \lambda_{max}^t \max_{i,j} \Big(\sum_{m=1}^n e_m(i)^2 \Big)^{1/2} \Big(\sum_{m=1}^n e_m(j)^2 \Big)^{1/2} && \text{by Cauchy-Schwarz} \\
&\leq \frac{1}{\min_k \pi(k)} \lambda_{max}^t && \text{See below}
\end{aligned}
$$

Note that $(\sum_{m=1}^n e_m(i)^2)^{1/2}$ is the norm of the ith row vector in the matrix E whose jth column is the vector e_j. $E^T E = I$ by orthonormality. The determinant of E is nonzero because its column vectors are linearly independent. Therefore, E is invertible, and $E^{-1} = E^T$. Therefore, $EE^T = I$. The (i,i)th entry EE^T equals 1, and it is the square of the norm of ith row vector of E. Therefore, $(\sum_{m=1}^n e_m(i)^2)^{1/2} = 1$. (Recall that $(\sum_{m=1}^n e_m(i)^2)^{1/2}$ is the norm of the ith row vector in of E.) Likewise, $(\sum_{m=1}^n e_m(j)^2)^{1/2} = 1$. QED

Note that the essence of the above proof is contained in the lemma. The rest is simple algebra with one fact about linear algebra at the end.

We now have a description of how fast P^t converges to Π. It converges at least as fast as λ_{max}^t. This will not help us much, however, as it is a painstaking process to find λ_{max} for arbitrary systems. Instead we shall focus on bounding λ_{max}. At first, we just want to bound λ_{max} away from 1, as this will guarantee convergence.

We will be able to find bounds for λ_{max} in terms of conductance when $\lambda_{max} = \lambda_2$. The reason is that the key to the bound involves relating λ_2 to the Dirichlet form, which in turn is related to conductance.

4 Random walks with $\lambda_{max} = \lambda_2$

Different random walks from the usual one can be given to guarantee that $\lambda_2 = \lambda_{max}$. Two examples work by shifting all the eigenvalues of the initial random walk to be non-negative. Then $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. So of course, $\lambda_{max} = \lambda_2$.

The first approach uses the following fact: If v is an eigenvector of V with eigenvalue λ, then v is an eigenvector of $P+I$ with eigenvalue $1+\lambda$. This is easy to see. $(I+P)v = Iv+Pv = v+\lambda v = (1+\lambda)v$. Therefore, all of the eigenvalues of P are between 2 and 0. Dividing $I+P$ by two gives a stochastic matrix, because the entries are all non-negative and the rows of $I+P$ sum to 2. $\frac{I+P}{2}v = \frac{1+\lambda}{2}v$. Therefore, all of the eigenvalues of $\frac{I+P}{2}$ are between 1 and 0. So $\lambda_2 = \lambda_{max}$ if our transition matrix is $\frac{I+P}{2}$. This corresponds to a random walk in which there is a probability of 1/2 of remaining at the same vertex.

The second approach turns the random walk into a continuous Markov chain by introducing a random jump time. This model is perhaps the more realistic of the two when it comes to crawling the Internet. Durret assumes that the jump times have a Poisson distribution. That is, the chance of making k jumps by time t is given by $e^{-t}\frac{t^k}{k!}$. This assumes a constant rate at which jumps occur an independence between the amount of time between jumps. To me, this sounds reasonable for a model of crawling the Internet. Let $H_t(x,y) = P(X_t = y|X_0 = x)$ where X_t is the state of the random walk (i.e. the position of the crawler) at time t. Let $J_t(k)$ be the event that exactly k jumps are made by time t.

$H_t(x,y) = \sum_{k=0}^{\infty} P(J_t(k))P(X_t = y|X_0 = x, J_t(k)) = \sum_{k=0}^{\infty} e^{-t}\frac{t^k}{k!}P^k(x,y)$

The eigenvalues of H_t are $e^{-t(1-\lambda_i)}$, $i = 1, 2, ..., n$. This is also easy to see. Let v be an eigenvector of P. Then $H_t v = \sum_{k=0}^{\infty} e^{-t}\frac{t^k}{k!}P^k(x,y)v = \sum_{k=0}^{\infty} e^{-t}\frac{t^k}{k!}\lambda^k v = e^{-t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}v = e^{-t}e^{\lambda t}v = e^{-t(1-\lambda)}v$. All the eigenvalues of H_t, then are positive, since the range of e^x is $(0,\infty)$. For H_t, $\lambda_{max} = e^{-t(1-\lambda_2)}$, because $e^{-t(1-\lambda)} = e^{-t}e^{\lambda t}$ is largest when λ is largest.

Even though the random walk is now continuous we still have the same bound as before for $\Delta(t)$. Durret gives a rather complicated proof of this result. It is much simpler to simply use the definition of H_t and the result for P^t. Note also that π is the stationary distribution of H, so we do not need to change π in the definition of $\Delta(t)$.

Proposition 8: $\Delta(t) \leq \frac{1}{\min_k \pi(k)}e^{-t(1-\lambda_2)} = \frac{1}{\min_k \pi(k)}\Lambda_{max}^t$ *where* $\Lambda_{max} = e^{-(1-\lambda_2)}$ *is* λ_{max} *for* H_1.

Proof:

$$
\begin{aligned}
\Delta(t) &= \max_{i,j} \Big| \frac{H_t(i,j)}{\pi(j)} - 1 \Big| \\
&= \max_{i,j} \Big| \frac{\sum_{k=0}^{\infty} e^{-t} \frac{t^k}{k!} P^k(x,y)}{\pi(j)} - 1 \Big| \\
&= \max_{i,j} \Big| \frac{\sum_{k=0}^{\infty} e^{-t} \frac{t^k}{k!} \Big(\pi(j) + \sqrt{\frac{\pi(j)}{\pi(i)}} \sum_{m=2}^{n} \lambda_m^k e_m(i) e_m(j) \Big)}{\pi(j)} - 1 \Big| && \text{by the lemma} \\
&= \max_{i,j} \Big| \frac{\sum_{k=0}^{\infty} e^{-t} \frac{t^k}{k!} \Big(\sqrt{\frac{\pi(j)}{\pi(i)}} \sum_{m=2}^{n} \lambda_m^k e_m(i) e_m(j) \Big)}{\pi(j)} \Big| \\
&= \max_{i,j} \Big| \frac{\sum_{k=0}^{\infty} e^{-t} \frac{t^k}{k!} \sum_{m=2}^{n} \lambda_m^k e_m(i) e_m(j)}{\sqrt{\pi(i)\pi(j)}} \Big| \\
&= \max_{i,j} \Big| \frac{\sum_{m=2}^{n} \sum_{k=0}^{\infty} e^{-t} \frac{t^k}{k!} \lambda_m^k e_m(i) e_m(j)}{\sqrt{\pi(i)\pi(j)}} \Big| \\
&= \max_{i,j} \Big| \frac{\sum_{m=2}^{n} \sum_{k=0}^{\infty} e^{-t} \frac{(\lambda_m t)^k}{k!} e_m(i) e_m(j)}{\sqrt{\pi(i)\pi(j)}} \Big| \\
&= \max_{i,j} \Big| \frac{\sum_{m=2}^{n} e^{-t} e^{\lambda_m t} e_m(i) e_m(j)}{\sqrt{\pi(i)\pi(j)}} \Big| \\
&= \max_{i,j} \Big| \frac{\sum_{m=2}^{n} e^{-t(1-\lambda_m)} e_m(i) e_m(j)}{\sqrt{\pi(i)\pi(j)}} \Big| \\
&\le \max_{i,j} \frac{\sum_{m=2}^{n} e^{-t(1-\lambda_m)} |e_m(i)||e_m(j)|}{\sqrt{\pi(i)\pi(j)}} \\
&\le \frac{\max_{i,j} \sum_{m=2}^{n} e^{-t(1-\lambda_m)} |e_m(i)||e_m(j)|}{\min_{i,j} \sqrt{\pi(i)\pi(j)}} \\
&\le \frac{\max_{i,j} \sum_{m=2}^{n} \max_m e^{-t(1-\lambda_m)} |e_m(i)||e_m(j)|}{\min_k \sqrt{\pi(k)\pi(k)}} \\
&= \frac{1}{\min_k \pi(k)} \max_m e^{-t(1-\lambda_m)} \max_{i,j} \sum_{m=2}^{n} |e_m(i)||e_m(j)| \\
&\le \frac{1}{\min_k \pi(k)} e^{-t(1-\lambda_2)} \max_{i,j} \sum_{m=1}^{n} |e_m(i)||e_m(j)| \\
&\le \frac{1}{\min_k \pi(k)} e^{-t(1-\lambda_2)} \max_{i,j} \Big(\sum_{m=1}^{n} e_m(i)^2 \Big)^{1/2} \Big(\sum_{m=1}^{n} e_m(j)^2 \Big)^{1/2} && \text{by Cauchy-Schwarz} \\
&\le \frac{1}{\min_k \pi(k)} e^{-t(1-\lambda_2)} && \text{as before}
\end{aligned}
$$

QED

With these examples finished, we shall turn to bounding $\lambda_2 = \lambda_{max}$ in terms of conductance. First, let us get a feel for the notion of conductance.

5 Conductance

Conductance measures the extent to which a graph is connected. Roughly speaking the easier it is to travel between any two vertices the betters the conductance. Following that notion, can also be considered to measure how hard it is to leave the surroundings of a vertex; the better the conductance the easier it is to leave.

Definition: Let G be a undirected multigraph with self-loops with vertex set V and edge set E.

1. $d_G(v)$ is the degree of a vertex, where self loops contribute 2 to the degree.
2. The volume of a set $S \subseteq V$ is $\text{vol}_G(S) = \sum_{v \in S} d_G(v)$
3. The cutset of $S \subseteq V$ is $C_G(S, S^C) = \{e \in E : \text{ one endpoint of } e \text{ is in } S \text{ and the other is in } S^C\}$
4. The conductance Φ_G of G is $\Phi_G = \min\left\{\frac{|C_G(S,S^C)|}{\text{vol}_G(S)} : \varnothing \neq S \subseteq V, \text{vol}_G(S) \leq \text{vol}_G(V)/2\right\}$.

I think it is best to define 0/0 as 0 for the conductance. This way any disconnected graph has conductance zero. Otherwise, graphs with one disconnected point added to a connected graph have nonzero conductance, but adding a self loop to the point gives zero conductance. Also this definition allows for graphs with $\text{vol}_G(V) < 2$. Finally, any graph then with more than one vertex has a conductance. A single vertex without a self loop has conductance 0, by this definition. I do not like this as the conductance dictates how quickly the invariant probability distribution is approached. And with only one state it is reached immediately. One was to get around this is to require $S \neq V$. This will be required in another essentially equivalent definition of conductance, and will not change anything except make the conductance of a single vertex undefined. I then propose that we define the conductance of a single vertex to be 1. As will be shown this is the largest possible conductance. To have one consistent definition we can define the conductance to be the minimum of the given set and the number 1. This will not change the conductance of any graph and will have the conductance of single vertices to be defined as 1.

Definition: $\Phi_G = \min\left\{1, \frac{|C_G(S,S^C)|}{\text{vol}_G(S)} : S \subseteq V, \varnothing \neq S \neq V, 0 \leq \text{vol}_G(S) \leq \text{vol}_G(V)/2\right\}$.

To show some of the bounds of conductance I will introduce the following definitions.

Definition: Let $S \subseteq V$ and $v \in S$.

1. The inner degree of a vertex v, $i_S(v)$, is the number edges that have v as one endpoint and the other endpoint in S, where self loops contribute 2 to the degree.
2. The outer degree of a vertex v, $o_S(v)$, is the number edges that have v as one endpoint and the other endpoint in S^C.

The following propositions follow clearly from these definitions.

Proposition 9:

1. $d_G(v) = i_S(v) + o_S(v)$

2. $|C_G(S, S^C)| = \sum_{v \in S} o_S(v)$

3. $|C_G(S, S^C)| \leq vol_G(S)$ *with equality iff* $i_S(v) = 0$ *for all* $v \in S$. *That is,* S *is a totally disconnected subgraph.*

4. $0 \leq \Phi_G \leq 1$.

5. $\Phi_G = 0$ *iff* G *is disconnected.*

Proof:

1. The degree of a vertex is the number of edges attached to it those can be partitioned into those whose other endpoint is in S and those that are in S^C. Note that the proposition holds because self loops contribute 2 to the degree in both i_S and in d_g.

2. $|C_G(S, S^C)|$ is the number of edges with one endpoint in S and one in S^C. This can be written as a sum over the vertices in S of the number of edges from that vertex to one in S^C.

3. $|C_G(S, S^C)| = \sum_{v \in S} o_S(v) \leq \sum_{v \in S}(i_S(v) + o_S(v)) = \sum_{v \in S} d_G(v) = \text{vol}_G(S)$. Equality holds iff $\sum_{v \in S} i_S(v) = 0$, which holds iff $i_S(v) = 0$ for all $v \in S$.

4. $0 \leq \frac{|C_G(S,S^C)|}{\text{vol}_G(S)} \leq 1$ by part 3. As Φ_G is the minimum of such element, $0 \leq \Phi_G \leq 1$.

5. Suppose $\Phi_G = 0$. Then there is a nonempty set S such that $\frac{|C_G(S,S^C)|}{\text{vol}_G(S)} = 0$. $|C_G(S, S^C)| = 0$. Therefore, there are no edges from S to S^C. So S and S^C separate G. Therefore, G is disconnected.

   Suppose instead that G is disconnected. Then there are disjoint nonempty sets $S, T \subseteq V$ such that $S \cup T = V$ and $C_G(S, T) = 0$. $\text{vol}_G(S) + \text{vol}_G(T) = \text{vol}_G(V)$. So $\text{vol}_G(S) \leq \text{vol}_G(V)/2$ or $\text{vol}_G(T) \leq \text{vol}_G(V)/2$. Without loss of generality, suppose $\text{vol}_G(S) \leq \text{vol}_G(V)/2$. Then $\frac{|C_G(S,S^C)|}{\text{vol}_G(S)} = \frac{|C_G(S,T)|}{\text{vol}_G(S)} = \frac{0}{\text{vol}_G(S)} = 0$. So Φ_G is the minimum of a set containing zero. Therefore, $\Phi_G \leq 0$. By part 4, $\Phi_G \geq 0$. So $\Phi_G = 0$.

QED

I was trying to understand the condition $\text{vol}_G(S) \leq \text{vol}_G(V)/2$. The following two facts will lead us to an equivalent definition of conductance from which it is easier to explain the condition. $\text{vol}_G(S) + \text{vol}_G(S^C) = \text{vol}_G(V)$. So for any set S, $\text{vol}_G(S) \leq \text{vol}_G(V)/2$ or $\text{vol}_G(S^C) \leq \text{vol}_G(V)/2$. Furthermore, $C_G(S, S^C) = C_G(S^C, S)$. Let $T \subseteq V$ be a set for which $\text{vol}_G(T) > \text{vol}_G(V)/2$. Let $S = T^C$. Then $\text{vol}_G(S) \leq \text{vol}_G(V)/2 < \text{vol}_G(T)$. So $\frac{|C_G(S,S^C)|}{\text{vol}_G(S)} = \frac{|C_G(T,T^C)|}{\min\{\text{vol}_G(T),\text{vol}_G(T^C)\}} = \frac{|C_G(S,S^C)|}{\min\{\text{vol}_G(S),\text{vol}_G(S^C)\}}$. Therefore, we may define the conductance as the minimum of $\frac{|C_G(S,S^C)|}{\min\{\text{vol}_G(S),\text{vol}_G(S^C)\}}$ over all $S \subseteq V$ with the exclusion of $\varnothing$ and V. The reason for excluding both $\varnothing$ and V is that then S or S^C equals $\varnothing$, and the term $0/0 = 0$ will always appear in the minimum, and as is clear from the definition conductance is always greater than or equal to zero.

Below is our new definition of conductance.

Definition: $\Phi_G = \min\left\{1, \frac{|C_G(S,S^C)|}{\min\{\text{vol}_G(S),\text{vol}_G(S^C)\}} : S \subseteq V, \varnothing \neq S \neq V\right\}$

The condition then that $\text{vol}_G(S) \leq \text{vol}_G(V)/2$ translates into having the minimum of $\text{vol}_G(S)$ and $\text{vol}_G(S^C)$. Why do we need this condition? Consider a case when $\text{vol}_G(S) \neq \text{vol}_G(S^C)$. Without loss of generality, assume $\mathrm{vol}_G(S^C) > \mathrm{vol}_G(S)$. Then $|C_G(S, S^C)| \leq \mathrm{vol}_G(S) < \mathrm{vol}_G(S^C)$. Therefore, $\frac{|C_G(S,S^C)|}{\mathrm{vol}_G(S^C)} < 1$. So if we don't have the minimum but rather choose the maximum or some other combination of $\mathrm{vol}_G(S)$ and $\mathrm{vol}_G(S^C)$, then the conductance will always be less than 1 (for any graph with more than 2 vertices). This ruins the beauty of having 1 being the ideal conductance for graphs.

Next let's get a sense of conductance by calculating it for several multigraphs. Then let's try and reason as to why the definition matches the heuristic notion of conductance described earlier. For calculations, I think it will be easier to use the second definition of conductance.

Example 1: *Let G be a mutligraph with n vertices and one edge connecting every vertex to every other vertex, with a self loop from every vertex to itself.*

Let $S \subseteq V$ with k vertices. $\mathrm{vol}_G(S) = (n+1)k$ and $\mathrm{vol}(V)_G = n(n+1)$. To have $\mathrm{vol}_G(S) \leq \mathrm{vol}_G(V)/2$, we need $k \leq n/2$. $|C_G(S, S^c)| = k(n-k)$. So $|C_G(S, S^c)|/\mathrm{vol}_G(S) = k(n-k)/((n+1)k) = (n-k)/(n+1)$. The minimum is achieved when k is as large as possible. For even n, the max k is $n/2$. In that case, $\Phi_G = n/(2(n+1))$. For odd $n \neq 1$, the max k is $(n-1)/2$. In that case, $\Phi_G = 1/2$.

Example 2: *Let G be a mutligraph with n vertices and (n-1) edges connecting one vertex to all the rest.*

There are two possible S to consider. S containing the hub and S not containing the hub. The degree of the hub is $n-1$ and $\mathrm{vol}_G(V) = 2(n-1)$. So if S contains the hub and $\mathrm{vol}_G(S) \leq \mathrm{vol}_G(V)/2$, then S is the set of just the hub. For such S, $|C_G(S, S^c)|/\mathrm{vol}_G(S) = (n-1)/(n-1) = 1$. For S not containing the hub but containing k vertices, $\mathrm{vol}_G(S) = k \leq (n-1) = \mathrm{vol}_G(V)/2$. So k can be anything from 1 to $n-1$. $|C_G(S, S^c)| = k$. Therefore, $|C_G(S, S^c)|/\mathrm{vol}_G(S) = k/k = 1$. So $\Phi_G = 1$.

The above examples show that the idea behind conductance is ease of getting from one place to another not just connectivity, because in the second example the graph has fewer connections but higher conductance. Related to this is the following easily verifiable fact: the conductance of a connected graph G is lower if a self loop is added to every vertex. This follows simply from the fact that $|C_G(S, S^c)| \neq 0$ will remain unchanged for every S while $\mathrm{vol}_G(S) \neq 0$ will increase.

Let's consider then a graph similar to that in example 1 but without self loops.

Example 3: *Let G be a mutligraph with n vertices and one edge connecting every vertex to every other vertex, with no self loop from every vertex to itself.*

Let $S \subseteq V$ with k vertices. $\mathrm{vol}_G(S) = (n-1)k$ and $\mathrm{vol}(V)_G = n(n-1)$. To have $\mathrm{vol}_G(S) \leq \mathrm{vol}_G(V)/2$, we need $k \leq n/2$. $|C_G(S, S^c)| = k(n-k)$. So $|C_G(S, S^c)|/\mathrm{vol}_G(S) = k(n-k)/((n-1)k) = (n-k)/(n-1)$. The minimum is achieved when k is as large as possible. For even n, the max k is $n/2$. In that case, $\Phi_G = n/(2(n-1))$. For odd $n \neq 1$, the max k is $(n-1)/2$. In that case, $\Phi_G = (n+1)/(2(n-1))$.

Finally, let's explore the effect of having multiple edges between vertices.

Example 4: *Let G be a multigraph with 2 vertices, a self loop on each vertex, and n edges connecting one vertex to the other.*

S must contain only one vertex for $\text{vol}_G(S) \leq \text{vol}_G(V)/2$. Then $|C_G(S, S^c)| = n$ and $\text{vol}_G(S) = n+2$. Therefore, $|C_G(S, S^c)|/\text{vol}_G(S) = n/(n+2)$. So more edges negate the effect of the self loops, which lower the conductance with $n = 1$ from 1 to 1/3.

Proposition 10: *Hubs are the only graphs with four or more vertices that have conductance equal to 1.*

Proof: Let G be a graph with 4 or more vertices such that $\Phi_G = 1$. Then G is a connected graph. For any set $S \neq \varnothing, V$ of vertices, $|C_G(S, S^c)| = \min\{\text{vol}_G(S), \text{vol}_G(S^c)\}$. If $\text{vol}_G(S) = |C_G(S, S^c)|$, then S has no internal edges; and if $\text{vol}_G(S^c) = |C_G(S, S^c)|$, then S^c has no internal edges. Since $|C_G(S, S^c)| = \min\{\text{vol}_G(S), \text{vol}_G(S^c)\}$, either S or S^c has no internal edges. Choose a set T of two vertices v_1, v_2 that are connected by at least one edge. Then T has an internal edge. Therefore, T^c has no internal edges. So every vertex in T^c has an edge connecting it to one of the two vertices in T. Therefore, v_1 or v_2 has an edge connecting it to a vertex in T^c. Without loss of generality, suppose v_1 is has an edge connecting it to a vertex $v_3 \in T^c$. Let T' be the set of v_1 and v_3. Then T' has an internal edge. So T'^c has no internal edges. Therefore, no vertex besides v_1 and v_3 can have an edge connecting it to v_2. Therefore, all of the edges connect v_1 to another vertex, except there may possibly be an edge connecting v_2 and v_3. Let v_4 be another vertex. Then there is an edge between v_1 and v_4. Let T'' be the set of v_1 and v_4. T'' has an internal edge. Therefore, T''^c has no internal edges. Since $v_2, v_3 \in T''^c$, there is no edge between v_2 and v_3. Therefore, all edges connect v_1 to another vertex. This type of graph is called a hub. QED

For graphs with fewer, than four vertices. There are ways to to have a conductance on 1 without having a hub.

Lastly, there is a way to relate conductance to the matrix $Q = DP$.

Proposition 11: $\Phi_G = \min\{1, \frac{Q(S,S^c)}{P(S)} : S \subseteq V, \varnothing \neq S \neq V, P(S) \leq 1/2\}$, *where* $Q(S, S^c) = \sum_{i \in S, j \in S^c} Q(i, j)$ *and* $P(S) = \sum_{i \in S} \pi(i)$.

Proof: $Q(S, S^c) = \sum_{i \in S, j \in S^c} Q(i, j) = \sum_{i \in S, j \in S^c} [DP](i, j) = \sum_{i \in S, j \in S^c} \pi(i) P(i, j) = \sum_{i \in S, j \in S^c} \frac{\deg(i)}{\text{vol}_G(V)} \frac{e(i,j)}{\deg(i)} = \sum_{i \in S, j \in S^c} \frac{e(i,j)}{\text{vol}_G(V)} = \frac{\sum_{i \in S, j \in S^c} e(i,j)}{\text{vol}_G(V)} = \frac{|C_G(S,S^c)|}{\text{vol}_G(V)}$, where $e(i, j)$ is the number of edges connecting the ith vertex and the jth vertex.

$P(S) = \sum_{i \in S} \pi(i) = \sum_{i \in S} \frac{\deg(i)}{\text{vol}_G(V)} = \frac{\sum_{i \in S} \deg(i)}{\text{vol}_G(V)} = \frac{\text{vol}_G(S)}{\text{vol}_G(V)}$.

Therefore, $\frac{Q(S,S^c)}{P(S)} = \frac{|C_G(S,S^c)|}{\text{vol}_G(S)}$, and the proposition is proved. QED

Note that $Q(i, j) = \frac{e(i,j)}{\text{vol}_G(V)}$ is the number of edges connecting the ith and jth vertices divided by twice the total number of vertices. Therefore, $\sum_{i,j} Q(i, j) = 1$. So Q may be viewed as a matrix of weights. We can therefore view a mulitgraph as a graph with only one edge between any two vertices with a given weight $Q(i, j)$ given to the edge. From this view point, the above proposition is a general definition of conductance.

6 Dirichlet form and Cheeger's inequality

The key fact to proving Cheeger's inequality is relating λ_2 to the Dirichlet form. Define an inner product on vectors of length n by $\langle x, y \rangle := \sum_{i=1}^{n} x(i)\pi(i)y(i)$.

Lemma 12: $\lambda_2 = \max\{\langle x, Px \rangle : E[x] = 0, E[x^2] = 1\}$

Proof: We will find the maximum of $\{\langle x, Px \rangle : E[x] = 0, E[x^2] = 1\}$ using Lagrange multipliers.

$\langle x, Px \rangle = \sum_{i=1}^{n}[x(i)\pi(i)\sum_{j=1}^{n} P(i,j)x(j)] = \sum_{i=1}^{n} x(i)\sum_{j=1}^{n}[DP](i,j)x(i) = \sum_{i=1}^{n} x(i)\sum_{j=1}^{n} Q(i,j)x(i)$. The constraint $E[x^2] = 1$ can be written as $\sum_{i=1}^{n}[x(i)^2\pi(i)] - 1 = 0$. Let $F(x,\lambda) = \langle x, Px \rangle - \lambda(\sum_{i=1}^{n}[\pi(i)x(i)^2] - 1)$.

Next we will take th partial derivative of F with respect to x_k and set it equal to zero. The resulting x are possible vectors that maximize F. Note that we are ignoring the constraint $E[x] = 0$ for the time being.

$$
\begin{aligned}
0 = \frac{\partial F}{\partial x(k)} &= \frac{\partial}{\partial x(k)}\Big[\sum_{i=1}^{n} x(i)\sum_{j=1}^{n} Q(i,j)x(j) - \lambda\Big(\sum_{i=1}^{n}[\pi(i)x(i)^2] - 1\Big)\Big] \\
&= \sum_{i\neq k}^{n} Q(i,k)x(i) + \sum_{j\neq k}^{n} Q(k,j)x(j) + 2Q(k,k)x(k) - 2\lambda\pi(k)x(k) \\
&= \sum_{i\neq k}^{n} Q(k,i)x(i) + \sum_{j\neq k}^{n} Q(k,j)x(j) + 2Q(k,k)x(k) - 2\lambda\pi(k)x(k) \\
&= 2\sum_{i\neq k}^{n} Q(k,i)x(i) + 2Q(k,k)x(k) - 2\lambda\pi(k)x(k) \\
&= 2\sum_{i=1}^{n} Q(k,i)x(i) - 2\lambda\pi(k)x(k) \\
&= 2\Big(\sum_{i=1}^{n} Q(k,i)x(i) - \lambda\pi(k)x(k)\Big)
\end{aligned}
$$

So $\sum_{i=1}^{n} Q(k,i)x(i) = \lambda\pi(k)x(k)$. $\sum_{i=1}^{n} Q(k,i)x(i)$ is the kth coordinate of $Qx.\pi(k)x(k)$ is the kth coordinate of Dx. Therefore, $Qx = \lambda Dx$. $Q = DP$ and is symmetric, so $Qx = Q^T x = P^T Dx$. So $P^T(Dx) = \lambda(Dx)$. Therefore, F is maximized when Dx is an eigenvector of P^T, and the maximum value is the eigenvalue λ. Therefore, $\max\{\langle x, Px \rangle : E[x] = 0, E[x^2] = 1\}$ is the largest eigenvalue of P^T whose corresponding eigenvector Dx satisfies $E[x] = 0$. The x, for which Dx are eigenvectors P^T, are orthogonal. ($\langle x_i, x_j \rangle = x_i^T D x_j = (D^{1/2}x_i)^T(D^{1/2}x_j) = 0$, because $D^{1/2}x$ is an eigenvalue of A.) The largest eigenvalue of P^T is 1; it corresponds to the eigenvector $(1, ..., 1)$, which does not have mean zero. For any other eigenvector x of P^T, $0 = \langle x, (1, ..., 1) \rangle = \sum_{i=1}^{n} x(i)\pi(i) = E[x]$. Therefore, the largest eigenvalue of P^T, whose corresponding eigenvector x has a mean zero, is the second largest eigenvalue of P^T, λ_2. So $\lambda_2 = \max\{\langle x, Px \rangle : E[x] = 0, E[x^2] = 1\}$. QED

The thought behind the theorem is just modifying the standard fact that for a symmetric matrix M, $x^T Mx$ is maximized under the constraint that x has a fixed norm when x is an eigenvector of M.

Next we define the Dirichlet form as follows: $\mathcal{E}(x, y) = \frac{1}{2}\sum_{i,j=1}^{n}(x(i) - x(j))(y(i) - y(j))Q(i, j)$. Another lemma relates the Dirichlet form to the inner product.

Lemma 13: $\mathcal{E}(x, x) = \langle x, x\rangle - \langle x, Px\rangle$

Proof:

$$
\begin{aligned}
\mathcal{E}(x, x) &= \frac{1}{2}\sum_{i,j=1}^{n}(x(i) - x(j))^2 Q(i, j) \\
&= \frac{1}{2}\Big[\sum_{i,j=1}^{n} x(i)^2 Q(i, j) + \sum_{i,j=1}^{n} x(j)^2 Q(i, j) - 2\sum_{i,j=1}^{n} x(i)x(j)Q(i, j)\Big] \\
&= \frac{1}{2}\Big[2\sum_{i,j=1}^{n} x(i)^2 Q(i, j) - 2\sum_{i,j=1}^{n} x(i)x(j)Q(i, j)\Big] \\
&= \sum_{i,j=1}^{n} x(i)^2 Q(i, j) - \sum_{i,j=1}^{n} x(i)x(j)Q(i, j) \\
&= \sum_{i,j=1}^{n} x(i)^2 \pi(i)P(i, j) - \sum_{i,j=1}^{n} x(i)x(j)Q(i, j) \\
&= \sum_{i=1}^{n} x(i)^2 \pi(i) - \sum_{i,j=1}^{n} x(i)x(j)Q(i, j) \\
&= \sum_{i=1}^{n} x(i)\pi(i)x(i) - \sum_{i,j=1}^{n} x(i)Q(i, j)x(j) \\
&= \sum_{i=1}^{n} x(i)\pi(i)x(i) - \sum_{i=1}^{n} x(i)\Big[\sum_{j=1}^{n} Q(i, j)x(j)\Big] \\
&= \sum_{i=1}^{n} x(i)\pi(i)x(i) - \sum_{i=1}^{n} x(i)\Big[\sum_{j=1}^{n} \pi(i)P(i, j)x(j)\Big] \\
&= \sum_{i=1}^{n} x(i)\pi(i)x(i) - \sum_{i=1}^{n} x(i)\pi(i)\Big[\sum_{j=1}^{n} P(i, j)x(j)\Big] \\
&= \sum_{i=1}^{n} x(i)\pi(i)x(i) - \sum_{i=1}^{n} x(i)\pi(i)[Px](i) \\
&= \langle x, x\rangle - \langle x, Px\rangle
\end{aligned}
$$

QED

Lemma 14: $\min\{\mathcal{E}(x, x) : E[x] = 0, \mathrm{Var}(x) = 1\} = 1 - \lambda_2$

Proof:

$$\begin{aligned}\min\{\mathcal{E}(x,x) : E[x]=0, \mathrm{Var}(x)=1\} &= \min\{\mathcal{E}(x,x) : E[x]=0, E[x^2]=1\}\\ &= \min\{\langle x,x\rangle - \langle x,Px\rangle : E[x]=0, E[x^2]=1\}\\ &= \min\{E[x^2] - \langle x,Px\rangle : E[x]=0, E[x^2]=1\}\\ &= \min\{1 - \langle x,Px\rangle : E[x]=0, E[x^2]=1\}\\ &= 1 - \max\{\langle x,Px\rangle : E[x]=0, E[x^2]=1\} = 1-\lambda_1\end{aligned}$$

QED

Theorem 15 (Cheeger's inequality)**:** $\lambda_2 < 1 - \frac{(\Phi_G)^2}{2}$

Proof: The idea of the theorem involves trying to establish an equality of the sort $\min\{\mathcal{E}(x,x) : E[x]=0, \mathrm{Var}(x)=1\} \geq f(\Phi_G)$ for some function f.

$$\begin{aligned}\min\{\mathcal{E}(x,x) : E[x]=0, \mathrm{Var}(x)=1\} &= \min\{\mathcal{E}(\frac{x}{\sqrt{\mathrm{Var}(x)}}, \frac{x}{\sqrt{\mathrm{Var}(x)}}) : E[x]=0\}\\ &= \min\left\{\frac{\mathcal{E}(x,x)}{\mathrm{Var}(x)} : E[x]=0\right\}\\ &= \min\left\{\frac{\mathcal{E}(x,x)}{\mathrm{Var}(x)} : x \text{ has median zero}\right\}\end{aligned}$$

The condition $E[x]=0$ could be replaced by the condition x has median zero, because this amount to shifting x by a constant, and $\mathcal{E}(x,x)$ and $\mathrm{Var}(x)$ remain unchanged if a constant is added to x. Our goal then is to show that $\mathcal{E}(x,x) \geq f(\Phi_G)\mathrm{Var}(x)$ for some function f for all x with median 0. $E[x^2] \geq \mathrm{Var}(x)$, so it is sufficient to show that $\mathcal{E}(x,x) \geq f(\Phi_G)E[x^2]$. We shall try to show this, because it is easier to work with $E[x^2]$ than $\mathrm{Var}(x)$. $\mathcal{E}(x,x) = \frac{1}{2}\sum_{i,j=1}^{n}(x(i)-x(j))^2 Q(i,j)$. So we need to show that $\frac{1}{2}\sum_{i,j=1}^{n}(x(i)-x(j))^2 Q(i,j) \geq f(\Phi_G)E[x^2]$. The last inequality will be denoted by $\star$.

At this point, it easier to begin with $\Phi_G E[x^2]$ and to try to bound this above by something related to $Q(i,j)$. An earlier proposition that related Φ_G and Q will be helpful for this purpose. It states $\Phi_G = \min\{1, \frac{Q(S,S^c)}{P(S)} : S \subseteq V, \varnothing \neq S \neq V, P(S) \leq 1/2\}$. So $\Phi(G) \leq \frac{Q(S,S^c)}{P(S)}$ for any S with $P(S) \leq 1/2$. Therefore, we shall want to write $E[x^2]$ in terms of a probability $P(S)$, which can be done using a well known fact in probability that $E[u] = \int_0^\infty P(u \geq t)dt$ for $u \geq 0$. Then $\Phi_G E[x^2]$ will be related to $Q(S,S^c)$. We will, however, need $P(S) \leq 1/2$. It is for this reason that we needed x to have median zero. And in evaluating $E[x^2]$, we will consider $x^2\mathrm{sgn}(x)$, which also has median zero.

$$
\begin{aligned}
\Phi_G E[x^2] &= \Phi_G E[|x^2 \mathrm{sgn}(x)|] \\
&= \Phi_G E[(x^2 \mathrm{sgn}(x))^+ + (x^2 \mathrm{sgn}(x))^-] \\
&= \Phi_G (E[(x^2 \mathrm{sgn}(x))^+] + E[(x^2 \mathrm{sgn}(x))^-]) \\
&= \Phi_G \Big(\int_0^\infty P((x^2 \mathrm{sgn}(x))^+ \geq t) dt + \int_0^\infty P((x^2 \mathrm{sgn}(x))^- \geq t) dt \Big) \\
&= \Phi_G \int_0^\infty P(\{i : x^2(i) \geq t, x(i) > 0\}) dt + \Phi_G \int_0^\infty P(\{i : x^2(i) \geq t, x(i) < 0\}) dt \\
&= \int_0^\infty \Phi_G P(S) dt + \int_0^\infty \Phi_G P(T) dt \text{ where } S = \{i : x^2(i) \geq t, x(i) > 0)\} \text{ and } T = \{i : x^2(i) \geq t, x(i) < 0)\} \\
&\leq \int_0^\infty \frac{Q(S, S^c)}{P(S)} P(S) dt + \int_0^\infty \frac{Q(T, T^c)}{P(T)} P(T) dt \qquad \text{because } P(S) \leq 1/2 \text{ and } P(T) \leq 1/2 \\
&= \int_0^\infty Q(S, S^c) dt + \int_0^\infty Q(T, T^c) dt \qquad \text{See footnote}^1 \\
&= \int_0^\infty \sum_{i \in S, j \in S^c} Q(i,j) dt + \int_0^\infty \sum_{i \in T, j \in T^c} Q(i,j) dt \\
&= \int_0^\infty \sum_{i,j} 1_S(i) 1_{S^c}(j) Q(i,j) dt + \int_0^\infty \sum_{i,j} 1_T(i) 1_{T^c}(j) Q(i,j) dt \\
&= \int_0^\infty \sum_{i,j} (1_S(i) 1_{S^c}(j) + 1_T(i) 1_{T^c}(j)) Q(i,j) dt \\
&\leq \sum_{i,j} \int_0^\infty 1_{(x^2(j), x^2(i)]}(t) Q(i,j) dt \qquad \text{because } x^2(i) \geq t > x^2(j) \Rightarrow i \in S, j \in S^c \text{ or } i \in T, j \in T^c \\
&= \sum_{i,j} Q(i,j) \int_0^\infty 1_{(x^2(j), x^2(i)]}(t) dt \\
&= \sum_{i,j : x^2(i) > x^2(j)} Q(i,j)(x^2(i) - x^2(j)) \\
&= \frac{1}{2} \sum_{i,j} Q(i,j) |x^2(i) - x^2(j)| \qquad \text{See footnote}^2 \\
&= \frac{1}{2} \sum_{i,j} |x(i) - x(j)| \sqrt{Q(i,j)} \cdot |x(i) + x(j)| \sqrt{Q(i,j)} \\
&\leq \frac{1}{2} \Big(\sum_{i,j} (x(i) - x(j))^2 Q(i,j) \Big)^{1/2} \Big(\sum_{i,j} (x(i) + x(j))^2 Q(i,j) \Big)^{1/2} \qquad \text{By Cauchy-Schwarz}
\end{aligned}
$$

The quantity $\big(\sum_{i,j} (x(i) - x(j))^2 Q(i,j)\big)^{1/2}$ needs to be squared at some point, and somehow the expression $\big(\sum_{i,j} (x(i) + x(j))^2 Q(i,j)\big)^{1/2}$ needs to go away. Preferably it will be related to $E[x^2]$ so that we may divide by it.

[1] There is potentially a problem for large enough t, because then S and T equal the empty set. But things work out because $Q(S, S^c)$ and $Q(T, T^c)$ are zero when S and T equal the emptyset.

[2] This calls for the Cauchy-Schwarz inequality to have it look like the quantity $\frac{1}{2} \sum_{i,j=1}^n (x(i) - x(j))^2 Q(i,j)$ on the right hand side of the inequality $\star$. First though we must factor the expression slightly.

$$\begin{aligned}
\Big(\sum_{i,j}(x(i)+x(j))^2Q(i,j)\Big)^{1/2} &= \Big(\sum_{i,j}(x^2(i)+x^2(j)+2x(i)x(j))Q(i,j)\Big)^{1/2} \\
&\le \Big(\sum_{i,j}(x^2(i)+x^2(j)+x^2(i)+x^2(j))Q(i,j)\Big)^{1/2} && \text{See footnote}^3 \\
&= \Big(2\sum_{i,j}x^2(i)Q(i,j)+2\sum_{i,j}x^2(j)Q(i,j)\Big)^{1/2} \\
&= \Big(2\sum_{i,j}x^2(i)Q(i,j)+2\sum_{i,j}x^2(j)Q(j,i)\Big)^{1/2} && \text{because } Q \text{ is symmetric} \\
&= \Big(4\sum_{i,j}x^2(i)Q(i,j)\Big)^{1/2} \\
&= 2\Big(\sum_{i,j}x^2(i)\pi(i)P(i,j)\Big)^{1/2} \\
&= 2\Big(\sum_{i}x^2(i)\pi(i)\Big)^{1/2} && \text{because } \sum_j P(i,j)=1 \\
&= 2(E[x^2])^{1/2}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Phi_G E[x^2] &\le \frac{1}{2}\Big(\sum_{i,j}(x(i)-x(j))^2Q(i,j)\Big)^{1/2}\Big(\sum_{i,j}(x(i)+x(j))^2Q(i,j)\Big)^{1/2} \\
&\le \Big(\sum_{i,j}(x(i)-x(j))^2Q(i,j)\Big)^{1/2}(E[x^2])^{1/2}
\end{aligned}$$

Squaring both sides and dividing by $2E[x^2]$ gets us $\frac{(\Phi_G)^2}{2}E[x^2] \le \frac{1}{2}\sum_{i,j}(x(i)-x(j))^2Q(i,j)$. That is the inequality $\star$ with $f(\Phi_G) = \frac{(\Phi_G)^2}{2}$. Thus, $1-\lambda_2 = \min\{\mathcal{E}(x,x) : E[x]=0, \mathrm{Var}(x)=1\} \ge \frac{(\Phi_G)^2}{2}$. So $\lambda_2 \le 1 - \frac{(\Phi_G)^2}{2}$. QED

Let's have a quick recap. We found that P^t converges to Π at a rate of at least λ_{max}^t. For some types of random walks, we have $\lambda_2 = \lambda_{max}$. Therefore, we want $\lambda_2 < 1$ to guarantee exponential convergence. Now that λ_2 is bounded in terms of the conductance. We can show that $\lambda_2 < 1$ by having $\Phi_G > 0$.

7 Theorem Bounding conductance for $G_{d,n}$

The article "On certain connectivity properties of the Internet topology" has one key theorem (theorem 1) bounding conductance away from zero for the graph $G_{d,n}$ as n approaches infinity.

[3] $0 \le (x(i)-x(j))^2 = x^2(i)+x^2(j)-2x(i)x(j)$. So $2x(i)x(j) \le x^2(i)+x^2(j)$.

First, let us describe the random graph $G_{d,n}$. The graph is formed by introducing what are called mini-vertices. Some of these mini-vertices are then identified to form one vertex in the final graph. The first mini-vertex is introduced with a single self-loop. Each additional mini-vertex added to the graph attaches to one of the previous mini-vertices with a probability proportional to the degree of the existing mini-vertices. This process continues until there are dn mini-vertices. The first d mini-vertices are then identified to form the first vertex; the next set of d mini-vertices are identified to form the second vertex; and so on.

Alternatively, you may view the graph as being built one vertex at a time with each vertex attaching with d edges to the other vertices with the possibility of self-loops.

The essence of theorem 1 in the paper can be split into two lemmas. The first bounds edge expansion. It is hinted in one sentence in the paper, but is not so obvious. I have included it next. The second is lemma 2 in the paper, which I have improved upon.

Lemma 16: *Supposed for fixed d, n, and an $\alpha \leq d$, $f(k) \geq P(A = Good(S))$ for all S with $|S| = k \leq n/2$ and for all A with $|A| = j$ for any $j \leq \alpha k - 1$. Then $P(\rho < \alpha) \leq \sum_{k=2}^{n/2} \binom{n}{k} \alpha k \binom{dn}{\alpha k} f(k)$. Note that αk may be real and that $\binom{dn}{\alpha k}$ is defined in terms of the gamma function.*

Definition: The edge expansion ρ_G of a graph G is given by
$\rho_G = \min\{\frac{|C_G(S,S^c)|}{|S|} : S \subseteq V, \varnothing \neq S \neq V, |S| \leq |V|/2\}$

We shall need the following definitions of good and bad mini-vertices. A mini-vertex is said to be associated with a set if its corresponding vertex is in the set.

Definition: A mini-vertex is good with respect to a set $S \subseteq V$ if it is associated with S and its father is associated with S^c or if it is associated with S^c and its father it associated with S. A mini-vertex is said to be bad with respect to S if it is not good with respect to S. By convention, mini-vertex 1 is bad. $Good(S)$ is the set of all mini-vertices that are good with respect to S.

Note that $|Good(S)| = |C_G(S, S^c)|$. Also note that every good mini-vertex contributes 1 to the volume of S; every bad mini-vertex associated with S contributes 2 to the volume of S; and every bad mini-vertex associated with S^c contributes 0 to the volume of S.

Proof:

$$
\begin{aligned}
P(\rho < \alpha) &= P\Big(\frac{|C_G(S,S^c)|}{|S|} < \alpha \text{ for some } S\Big) && \text{where } S \subseteq V \text{ and } |S| \le |V|/2 \\
&= P\Big(\bigcup_S \Big\{\frac{|C_G(S,S^c)|}{|S|} < \alpha\Big\}\Big) \\
&= P\Big(\bigcup_{k=1}^{n/2} \bigcup_{S:|S|=k} \Big\{\frac{|C_G(S,S^c)|}{|S|} < \alpha\Big\}\Big) \\
&= P\Big(\bigcup_{k=1}^{n/2} \bigcup_{S:|S|=k} \Big\{|C_G(S,S^c)| < \alpha k\Big\}\Big) \\
&= P\Big(\bigcup_{k=2}^{n/2} \bigcup_{S:|S|=k} \Big\{|C_G(S,S^c)| < \alpha k\Big\}\Big) && \text{If } k=1, |C_G(S,S^c)| \ge d \ge \alpha = \alpha k \\
&= P\Big(\bigcup_{k=2}^{n/2} \bigcup_{S:|S|=k} \Big\{|Good(S)| < \alpha k\Big\}\Big) \\
&= P\Big(\bigcup_{k=2}^{n/2} \bigcup_{S:|S|=k} \bigcup_{j=0}^{\lfloor \alpha k-1 \rfloor} \Big\{|Good(S)| = j\Big\}\Big) \\
&= P\Big(\bigcup_{k=2}^{n/2} \bigcup_{S:|S|=k} \bigcup_{j=0}^{\lfloor \alpha k-1 \rfloor} \Big\{Good(S) = A \text{ for an } A \text{ with } |A| = j\Big\}\Big) && \text{where } A \text{ is a subset of mini-vertices} \\
&= P\Big(\bigcup_{k=2}^{n/2} \bigcup_{S:|S|=k} \bigcup_{j=0}^{\lfloor \alpha k-1 \rfloor} \bigcup_{A:|A|=j} \Big\{A = Good(S)\Big\}\Big) \\
&\le \sum_{k=2}^{n/2} \sum_{S:|S|=k} \sum_{j=0}^{\lfloor \alpha k-1 \rfloor} \sum_{A:|A|=j} P(A = Good(S)) \\
&\le \sum_{k=2}^{n/2} \sum_{S:|S|=k} \sum_{j=0}^{\lfloor \alpha k-1 \rfloor} \sum_{A:|A|=j} f(k) \\
&\le \sum_{k=2}^{n/2} \sum_{S:|S|=k} \sum_{j=0}^{\lfloor \alpha k-1 \rfloor} \binom{dn}{j} f(k) && \text{there are } \binom{dn}{j} \text{ ways to choose } A \\
&\le \sum_{k=2}^{n/2} \sum_{S:|S|=k} \sum_{j=0}^{\lfloor \alpha k-1 \rfloor} \binom{dn}{\alpha k} f(k) && \text{since } \binom{dn}{j} \text{ increases until } j = dn/2 \\
&\le \sum_{k=2}^{n/2} \sum_{S:|S|=k} \alpha k \binom{dn}{\alpha k} f(k) \\
&= \sum_{k=2}^{n/2} \binom{n}{k} \alpha k \binom{dn}{\alpha k} f(k) && \text{there are } \binom{n}{k} \text{ ways to choose } S
\end{aligned}
$$

QED

I shall next work to improve lemma two presented in the paper. The lemma works to bound $P(A = Good(S))$

Lemma 17: $P(A = Good(S)) \leq \binom{dn}{|A|/2} \times \binom{dn-|A|/2}{dk-|A|/2}^{-1}$

Proof: $P(A = Good(S))$ is calculated mini-vertex by mini-vertex as follows. Consider the graph at time $j \in \{0, 1, ..., dn\}$ at each time one mini-vertex is added. Let $A^j = A \cap \{1, ..., j\}$ and let $S^j = S_1 \cap \{0, 1, ..., j\}$ where S_1 is the set of mini-vertices associated with S. Then

$$\begin{aligned}
P(A = Good(S)) &= \Big[\prod_{j=1}^{dn} P(A^j = Good(S^j)|A^{j-1} = Good(S^{j-1}))\Big] P(A^0 = Good(S^0)) \\
&= \Big[\prod_{j=1}^{dn} P(A^j = Good(S^j)|A^{j-1} = Good(S^{j-1}))\Big] \qquad \text{because} A^0 = S^0 = \varnothing \\
&= \prod_{j \notin A} P(A^j = Good(S_j)|A^{j-1} = Good(S^{j-1})) \prod_{j \in A} P(A^j = Good(S^j)|A^{j-1} = Good(S^{j-1})) \\
&\leq \prod_{j \notin A} P(A^j = Good(S_j)|A^{j-1} = Good(S^{j-1})) \\
&= \prod_{j \notin A} P(j \text{ is bad}|A^{j-1} = Good(S^{j-1}))
\end{aligned}$$

The previous inequality was not mentioned in the paper, which could create some confusion as they use A to denote A^j for all times. On another note, I think it might be possible to get a better bound by including the product over $j \in A$.

Let's proceed to bound $P(j \text{ is bad}|A^{j-1} = Good(S^{j-1}))$ for $j \notin A$. To begin, let us name the mini-vertices in A^c. The mini-vertices are named differently based upon whether it is associated with S or S^c. This is done to produce a smaller lower bound as I found out the hard way.

Note that $|S| = k$. Let S_1 be the set of mini-vertices associated with S, and let S_2 be the set of mini-vertices associated with S^c. Let $A_1 = A \cap S_1$ and $A_2 = A \cap S_2$. Let $k_1 = |A_1|$ and $k_2 = |A_2|$. Then $|A| = k_1 + k_2$. The mini-vertices in A^c are either in $S_1 \setminus A_1$ or in $S_2 \setminus A_2$. $|S_1 \setminus A_1| = dk - k_1$, because there are dk mini-vertices associated with S. That is, there are $dk - k_1$ mini-vertices in $S_1 \setminus A_1$. Denote them by $x_1 < x_2 < \cdots < x_{dk-k_1}$. Note that x_i is just a number. It's the (x_i)th mini-vertex added to the graph. For each x_i, let y_i be the number of mini-vertices that arrived prior to x_i and belong to A and let z_i be the total number of mini-vertices that arrived prior to x_i and belong to A^c. Then $y_i + z_i$ is the total number of mini-vertices that arrive before x_i. So $x_i = y_i + z_i + 1$. $|S_2 \setminus A_2| = dn - dk - k_2$, because there are $dn - dk$ mini-vertices associated with S^c. That is, there are $dn - dk - k_2$ mini-vertices in $S_2 \setminus A_2$. Denote them by $\bar{x}_1 < \bar{x}_2 < \cdots < \bar{x}_{dn-dk-k_2}$. Note that $\bar{x}_i$ is just a number. It's the $(\bar{x}_i)$th mini-vertex added to the graph. For each $\bar{x}_i$, let $\bar{y}_i$ be the number of mini-vertices that arrived prior to $\bar{x}_i$ and belong to A and let $\bar{z}_i$ be the total number of mini-vertices that arrived prior to $\bar{x}_i$ and belong to A^c. Then $\bar{y}_i + \bar{z}_i$ is the total number of mini-vertices that arrive before $\bar{x}_i$. So $\bar{x}_i = \bar{y}_i + \bar{z}_i + 1$.

Next I will show $P(x_i \text{ is bad}|A^{x_i-1} = Good(S^{x_i-1})) \leq \frac{2(i-1)+y_i}{2(y_i+z_i)}$ and $P(\bar{x}_i \text{ is bad}|A^{\bar{x}_i-1} = Good(S^{\bar{x}_i-1})) \leq \frac{2(i-1)+\bar{y}_i}{2(\bar{y}_i+\bar{z}_i)}$. The proof for x_i and $\bar{x}_i$ is essentially the same so I shall only prove it for x_i.

x_i is bad iff it attaches to a mini-vertex in S_1. Therefore, the probability that x_i is bad is the probability that x_i attaches to a mini-vertex in S_1, which is the volume of S_1 just before x_i arrives divided by the volume of the graph just before x_i arrives. This is the preferential attachment model that was given.

Just before x_i arrives, there are $y_i + z_i$ mini-vertices in the graph; each one contributes 2 to the volume of the graph. Therefore, the volume of the graph just before x_i arrives is $2(y_i + z_i)$.

To calculate the volume of S_1 just before x_i arrives, we can split the mini-vertices into two types: bad mini-vertices and good ones. The bad mini-vertices that contribute to the volume of S_1 are in $S_1 \setminus A_1$. The bad mini-vertices in S_2 do not contribute to the volume. The good mini-vertices are all in A, because we are calculating the probability given that A contains no bad mini-vertices just before x_i arrives.

Just before x_i arrived, the vertices in $S_1 \setminus A_1$, were $x_1, ..., x_{i-1}$. There are $i-1$ such vertices. Therefore there are at most $i-1$ bad mini-vertices that contribute to the volume of S_1. Each contributes at most 2 to the volume of S_1. y_i denotes the numbers of mini-vertices in A just before x_i arrives. Since we are calculating the probability given that A contains no bad mini-vertices just before x_i arrives, these y_i mini-vertices are all good. So each one contributes 1 to the volume of S_1. Therefore, the volume of S_i just before x_i arrives is at most $2(i-1) + y_i$.

So $P(x_i \text{ is bad}|A^{x_i-1} = Good(S^{x_i-1})) \leq \frac{2(i-1)+y_i}{2(y_i+z_i)}$.

Next, the paper finds a larger bound that is easier to work with. $\frac{2(i-1)+y_i}{2(y_i+z_i)} \leq \frac{i+|A|}{z_i+1+|A|}$ and $\frac{2(i-1)+\bar{y}_i}{2(\bar{y}_i+\bar{z}_i)} \leq \frac{i+|A|}{\bar{z}_i+1+|A|}$. I have managed to improve these bounds. $\frac{2(i-1)+y_i}{2(y_i+z_i)} \leq \frac{i+|A|/2}{z_i+1+|A|/2}$ and $\frac{2(i-1)+\bar{y}_i}{2(\bar{y}_i+\bar{z}_i)} \leq \frac{i+|A|/2}{\bar{z}_i+1+|A|/2}$. Again the proofs for the two inequalities are identical, so I shall only show the first.

The key fact they use is that $\frac{a+x}{b+x}$ is an increasing function of x for $b > a > 0$. Also note that z_i is all of the bad mini-vertices there are by time x_i. The $x_1, ..., x_{i-1}$ are some of these bad vertices. So $z_i \geq i-1$. In addition, $|A| \geq y_i$ because y_i is the number of mini-vertices in A by a given time.

$$\begin{aligned}
\frac{2(i-1)+y_i}{2(y_i+z_i)} &\leq \frac{2(i-1)+y_i}{2z_i+y_i} && \text{Subtract } y_i \text{ from the denominator} \\
&\leq \frac{(i-1)+y_i/2}{z_i+y_i/2} \\
&\leq \frac{i+y_i/2}{z_i+1+y_i/2} && \text{Adding 1 to the numerator and denominator} \\
&\leq \frac{i+|A|/2}{z_i+1+|A|/2} && \text{because } |A| \geq y_i
\end{aligned}$$

Note that as we range over all bad mini-vertices x_i and $\bar{x}_i$, the number of bad mini-vertices that have arrived before x_i and $\bar{x}_i$ range over the set $\{0, 1, ..., dn-|A|-1\}$, because there are $dn-|A|$ bad mini-vertex arrives and when the last bad mini-vertex arrives the number of bad mini-vertices that have arrived beforehand is $dn-|A|-1$. Therefore, $\bigcup_{i=1}^{dk-k_1}\{z_i\}\cup\bigcup_{i=1}^{dn-dk-k_2}\{\bar{z}_i\} = \{0,1,2,...,dn-|A|-1\}$. Or as we shall need, $\bigcup_{i=1}^{dk-k_1}\{z_i+1\}\cup\bigcup_{i=1}^{dn-dk-k_2}\{\bar{z}_i+1\} = \{1,2,...,dn-|A|\}$. Call this $\star\star$.

Below is a summary of the key inequalities so far.

$$
\begin{aligned}
P(A = Good(S)) &\leq \prod_{j\notin A} P(j \text{ is bad}|A^{j-1} = Good(S^{j-1})) \\
&\leq \prod_{i=1}^{dk-k_1} \frac{2(i-1)+y_i}{2(y_i+z_i)} \prod_{i=1}^{dn-dk-k_2} \frac{2(i-1)+\bar{y}_i}{2(\bar{y}_i+\bar{z}_i)} \\
&\leq \prod_{i=1}^{dk-k_1} \frac{i+|A|}{z_i+1+|A|} \prod_{i=1}^{dn-dk-k_2} \frac{i+|A|}{\bar{z}_i+1+|A|}
\end{aligned}
$$

$$
\begin{aligned}
\prod_{i=1}^{dk-k_1} \frac{i+|A|/2}{z_i+1+|A|/2} \prod_{i=1}^{dn-dk-k_2} \frac{i+|A|/2}{\bar{z}_i+1+|A|/2} &= \frac{\prod_{i=1}^{dk-k_1}(i+|A|/2)\prod_{i=1}^{dn-dk-k_2}(i+|A|/2)}{\prod_{i=1}^{dk-k_1}(z_i+1+|A|/2)\prod_{i=1}^{dn-dk-k_2}(\bar{z}_i+1+|A|/2)} \\
&= \frac{\prod_{i=1}^{dk-k_1}(i+|A|/2)\prod_{i=1}^{dn-dk-k_2}(i+|A|/2)}{\prod_{i=1}^{dn-|A|}(i+|A|/2)} \text{ by } \star\star \\
&= \frac{(|A|/2)!\Big[\prod_{i=1}^{dk-k_1}(i+|A|/2)\Big](|A|/2)!\Big[\prod_{i=1}^{dn-dk-k_2}(i+|A|/2)\Big]}{(|A|/2)!\Big[\prod_{i=1}^{dn-|A|}(i+|A|/2)\Big](|A|/2)!} \\
&= \frac{(dk-k_1+|A|/2)!(dn-dk-k_2+|A|/2)!}{(dn-|A|/2)!(|A|/2)!} \\
&= \frac{(dk-k_1/2+k_2/2)!(dn-dk-k_2/2+k_1/2)!}{(dn-|A|/2)!(|A|/2)!} \\
&= \frac{(dn)!}{(dn-|A|/2)!(|A|/2)!}\frac{(dk-k_1/2+k_2/2)!(dn-dk-k_2/2+k_1/2)!}{(dn)!} \\
&= \binom{dn}{|A|/2} \times \binom{dn}{dk-k_1/2+k_2/2}^{-1} \\
&\leq \binom{dn}{|A|/2} \times \binom{dn}{dk-|A|/2}^{-1} \text{ Dividing by a smaller number} \\
&\leq \binom{dn}{|A|/2} \times \binom{dn-|A|/2}{dk-|A|/2}^{-1} \text{ Dividing by smaller number}
\end{aligned}
$$

QED

Theorem 18: *For all integers $d \geq 2$, $\lim_{n\to\infty} P(\Phi_{G_{d,n}} < \frac{\alpha}{2d+\alpha}) = 0$ for all $\alpha < \min\{\frac{2d-3}{4}, \alpha_0\}$ where α_0 is the unique positive solution less than d of the equation $\frac{3}{2}\alpha(1+\ln(d)-\ln(\alpha)) = (d-1-\frac{5}{2}\alpha)\ln 2$.*

Proof: The idea of the proof is to first bound Φ_G below in terms of ρ_G. Then the lemmas shown that $\lim_{n\to\infty} P(\rho_{G_{d,n}} < \alpha) = 0$.

For any $S \subseteq V$, $\text{vol}_G(S)$ is contributed by internal edges and edges from S^c to S. There are at most $d|S|$ internal edges, each of which contributes 2 to the volume of S. There are $|C_G(S,S^c)|$ edges from S^c to S, each of which contributes 1 to the volume of S. Thus, $\text{vol}_G(S) \leq 2d|S| + |C_G(S,S^c)|$. Therefore, $\min\{\text{vol}_G(S), \text{vol}_G(S^c)\} \leq 2d\min\{|S|,|S^c|\} + |C_G(S,S^c)|$.

$$\begin{aligned}
\Phi_G &= \min\left\{1, \frac{|C_G(S,S^C)|}{\min\{\text{vol}_G(S),\text{vol}_G(S^c)\}} : S\subseteq V, \varnothing \neq S \neq V\right\} \\
&\geq \min\left\{\frac{|C_G(S,S^C)|}{2d\min\{|S|,|S^c|\} + |C_G(S,S^c)|} : S\subseteq V, \varnothing \neq S \neq V\right\} \\
&= \min\left\{\frac{|C_G(S,S^C)|}{2d|S| + |C_G(S,S^c)|} : S\subseteq V, \varnothing \neq S \neq V, |S| \leq |V|/2\right\} \\
&= \min\left\{\frac{\frac{|C_G(S,S^C)|}{|S|}}{2d + \frac{|C_G(S,S^c)|}{|S|}} : S\subseteq V, \varnothing \neq S \neq V, |S| \leq |V|/2\right\} \\
&\geq \frac{\min\left\{\frac{|C_G(S,S^C)|}{|S|} : S\subseteq V, \varnothing \neq S \neq V, |S| \leq |V|/2\right\}}{2d + \min\left\{\frac{|C_G(S,S^c)|}{|S|} : S\subseteq V, \varnothing \neq S \neq V, |S| \leq |V|/2\right\}} \\
&= \frac{\rho_G}{2d+\rho_G}
\end{aligned}$$

Then $P(\Phi_{G_{d,n}} < \frac{\alpha}{2d+\alpha}) \leq P(\rho_{G_{d,n}} < \alpha)$.
So if $\lim_{n\to\infty} P(\rho_{G_{d,n}} < \alpha) = 0$, then $\lim_{n\to\infty} P(\Phi_{G_{d,n}} < \frac{\alpha}{2d+\alpha}) = 0$.

Combining the last two lemmas we have that $P(\rho_{G_{d,n}} < \alpha) \leq \sum_{k=2}^{n/2} \binom{n}{k}\alpha k\binom{dn}{\alpha k}\binom{dn}{|A|/2} \times \binom{dn-|A|/2}{dk-|A|/2}^{-1}$. Following several inequalities presented in the paper involving factorials and $\binom{n}{k}$ and modifying one slightly, I found that $P(\rho_{G_{d,n}} < \alpha) \leq \sum_{k=2}^{n/2} \alpha k(2)^{\frac{\alpha}{2}k}(\frac{ed}{\alpha})^{\frac{3}{2}\alpha k}(\frac{k}{n})^{(d-1-2\alpha)k}$.

The paper then finds a neat way to bound the sum. They show that the largest term in the sum is either the first term when $k=2$ or the last term when $k=n/2$. Let $f(k) = \alpha k(2)^{\frac{\alpha}{2}k}(\frac{ed}{\alpha})^{\frac{3}{2}\alpha k}(\frac{k}{n})^{(d-1-2\alpha)k}$. $f'(k) = \alpha(2)^{\frac{\alpha}{2}k}(\frac{ed}{\alpha})^{\frac{3}{2}\alpha k}(\frac{k}{n})^{(d-1-2\alpha)k}\Big[1 + k(\frac{\alpha}{2}\ln 2 + \frac{3}{2}\alpha(1+\ln\frac{d}{\alpha}) + (d-1-2\alpha)(\ln\frac{k}{n}+1)\Big]$. The sign of f' is determined by the sign of $g(k) = 1 + k(\frac{\alpha}{2}\ln 2 + \frac{3}{2}\alpha(1+\ln\frac{d}{\alpha}) + (d-1-2\alpha)(\ln\frac{k}{n}+1)$. $g''(k) = \frac{d-1-2\alpha}{k} > 0$ if $\alpha < \frac{d-1}{2}$. That is, g is concave up. It can also be shown that $g(2) < 0$ for large enough n and $g(n/2) > 0$. Therefore, g is negative up until some point in the interval $(2,n/2)$. Past that point g is positive. Remember that g and f' have the same sign. Therefore, $f(k)$ is decreasing up until some point in the interval $(2,n/2)$. Past that point $f(k)$ is increasing. Therefore, $f(k)$ attains its maximum on the interval $[2,n/2]$ at 2 or $n/2$.

Suppose it attains its maximum at $k = 2$. Then
$P(\rho_{G_{d,n}} < \alpha) \leq \sum_{k=2}^{n/2} f(2) \leq \frac{n}{2} f(2) = n\alpha(2)^{\alpha}(\frac{ed}{\alpha})^{3\alpha}(\frac{2}{n})^{2(d-1-2\alpha)} = \alpha(2)^{\alpha}(\frac{ed}{\alpha})^{3\alpha}2^{2(d-1-2\alpha)}(\frac{1}{n})^{2(d-1-2\alpha)-1}$, which goes to zero as n goes to infinity if $2(d-1-2\alpha)-1 > 0$. So α must satisfy $\alpha < \frac{2d-3}{4}$. If α satisfies this condition, then α already satisfies the condition $\alpha < \frac{d-1}{2}$ that was mentioned earlier.

If $f(k)$ attains its maximum at $k = n/2$. Then
$P(\rho_{G_{d,n}} < \alpha) \leq \sum_{k=2}^{n/2} f(n/2) \leq \frac{n}{2} f(n/2) = \frac{n}{2}\alpha\frac{n}{2}(2)^{\frac{\alpha n}{4}}(\frac{ed}{\alpha})^{\frac{3}{4}\alpha n}(\frac{1}{2})^{(d-1-2\alpha)\frac{n}{2}} = \frac{\alpha}{4}n^2\left[(2)^{\frac{\alpha}{2}}(\frac{ed}{\alpha})^{\frac{3}{2}\alpha}(\frac{1}{2})^{(d-1-2\alpha)}\right]^{\frac{n}{2}}$, which goes to zero as n approaches infinity if $(2)^{\frac{\alpha}{2}}(\frac{ed}{\alpha})^{\frac{3}{2}\alpha}(\frac{1}{2})^{(d-1-2\alpha)} < 1$. That inequality is equivalent to $(\frac{ed}{\alpha})^{\frac{3}{2}\alpha} < 2^{(d-1-\frac{5}{2}\alpha)}$. Taking the natural logarithm of both sides we find that α must satisfy $\frac{3}{2}\alpha(1+\ln(d)-\ln(\alpha)) < (d-1-\frac{5}{2}\alpha)\ln 2$. For α near zero this inequality holds. The left side of the inequality is monotonically increasing with α for $\alpha < d$ ($\alpha < d$ by the earlier condition.), and the right of the inequality is monotonically decreasing with α. Therefore, there is a largest α_0 such that for all $\alpha < \alpha_0$, the inequality holds.

So for all $\alpha < \min\{\frac{2d-3}{4}, \alpha_0\}$, $\lim_{n\to\infty} P(\rho_{G_{d,n}} < \alpha) = 0$. Therefore, $\lim_{n\to\infty} P(\Phi_{G_{d,n}} < \frac{\alpha}{2d+\alpha}) = 0$.
QED

Below are approximations of the supremum of all possible α for $d = 2, 3, 4, 5$ using my version of theorem 1 and the one given in the paper.

$\alpha <$	d=2	d=3	d=4	d=5
Their bound	0.01034491	0.13298409	0.20000000	0.20000000
My bound	0.08743082	0.18755789	0.29006993	0.39349522

In the following table are the corresponding values of $\frac{\alpha}{2d+\alpha}$ that the conductance is bounded above.

$\Phi_G \geq$	d=2	d=3	d=4	d=5
Their bound	0.00257956	0.02168342	0.02439024	0.01960784
My bound	0.02139017	0.03031210	0.03499005	0.03785976

Using Cheeger's inequality and the above numbers, we can find the values that λ_2 is bounded below. They are shown in the table below.

$\lambda_2 \leq$	d=2	d=3	d=4	d=5
Their bound	0.99999667	0.99976491	0.99970256	0.99980777
My bound	0.99977123	0.99954059	0.99938785	0.99928332

As can be seen these are not great bounds for λ_2. Even with my improved bounds λ_2 could still be close to 1. We conjectured that maybe λ_2 approaches a constant as n approaches infinity. I ran a program to test this, and the results support this conjecture. Assuming that λ_2 indeed approaches a constant, the table below shows estimates of these values that I found by running the program 100 times with 2000 vertices.

λ_2	d=2	d=3	d=4	d=5
	0.8354	0.7212	0.6428	0.5852

I conjecture that the constant λ_2 approaches decreases with d. In addition, I conjecture that $\lambda_2 \geq 0.5$ for all d.

Lastly, I managed to prove that the result of the theorem does not hold when $d = 1$.

Theorem 19: *For* $d = 1$, $\lim_{n\to\infty} P(\Phi_{G_{d,n}} > \beta) = 0$ *for all* $\beta > 0$.

Proof: Let S_n be the set of the second vertex added to the graph and all of its descendants at time n. Then S_n^c is the set of the first vertex and all of its descendants except for the second vertex and its descendants. $|C_G(S_n, S_n^c)| = 1$ for all n because the only edge in $C_G(S_n, S_n^c)$ is the edge connecting the first and second vertices.

Therefore, $\Phi_{G_{d,n}} \leq \frac{|C_G(S_n,S_n^c)|}{\min\{\mathrm{vol}_{G_{d,n}}(S_n),\mathrm{vol}_{G_{d,n}}(S_n^c)\}} = \frac{1}{\min\{\mathrm{vol}_{G_{d,n}}(S_n),\mathrm{vol}_{G_{d,n}}(S_n^c)\}}$

For all $\beta > 0$, $P(\Phi_{G_{d,n}} > \beta) \leq P(\min\{\mathrm{vol}_{G_{d,n}}(S_n), \mathrm{vol}_{G_{d,n}}(S_n^c)\} < \frac{1}{\beta})$. It is a standard result for $d = 1$ that $\lim_{n\to\infty} P(\min\{\mathrm{vol}_{G_{d,n}}(S_n), \mathrm{vol}_{G_{d,n}}(S_n^c)\} < \frac{1}{\beta}) = 0$. Thus, $\lim_{n\to\infty} P(\Phi_{G_{d,n}} > \beta) = 0$ for all $\beta > 0$. QED