The self-avoiding walk-jump (SAWJ) algorithm for finding maximum degree nodes in large graphs

Jonathan Stokes and Steven Weber
Department of Electrical and Computer Engineering
Drexel University, Philadelphia, PA, USA

***Abstract*—In this paper we present a self-avoiding walk-jump (SAWJ) algorithm for finding a maximum degree node on a large assortative graph. We offer two contributions: i) we use the theory of absorbing Markov chains to effectively approximate the required search time as a function of the number of nodes, the edge density, and the assortativity, and ii) we measure the performance of our algorithm against competing algorithms (including star sampling) from the literature on the class of assortative Erdős-Rényi (AER) random graphs, and on six real-world large graphs from the Stanford SNAP dataset. In most cases SAWJ significantly outperforms the competing algorithms.**

***Index Terms*—social networks; graph search; random graphs; random walks; random sampling; graph rewiring; assortativity.**

I. Introduction

A. Motivation

Many large datasets are in the form of *graphs*, e.g., i) social networks like Facebook or LinkedIn, ii) community structures like arXiv paper authorship or IMDb, or iii) system diagrams like protein networks or the electrical grid. In graph's edges denote influence or connection among nodes, hence maximum degree nodes are, in a sense, the most influential or most highly connected. It is a natural objective to seek out maximum degree nodes, and for large graphs this must be done by a search or sampling algorithm. We emphasize the distinction here between finding (or estimating) the maximum degree and finding a maximum (or near-maximum) degree node. We focus here on actually finding a max degree node.

In this paper, we *assume* the maximum degree, denoted δ, is known *a priori*, and so the problem is to find any member of the set of maximum degree nodes, denoted $\mathcal{V}_\delta$. Although this assumption appears artificial, it is not substantially different from the more practical scenario where δ is unknown and the time available for the search is bounded. That is, the same algorithm may well be used to i) find the largest degree possible, without knowledge of δ, over a bounded search time, and ii) knowing δ, seek to find any node in $\mathcal{V}_\delta$, in as short a time as possible.

It is also natural to consider the problem of finding *all* the maximum degree nodes, i.e., the entire set $\mathcal{V}_\delta$, instead of just any one member of $\mathcal{V}_\delta$. Interestingly, as shown in our prior work [1], for an important class of random graphs, with high probability there is a unique maximum degree node, i.e., $|\mathcal{V}_\delta| \approx 1$. Thus algorithms for our problem of finding one maximizer will also apply to the problem of finding all.

Without knowing the graph structure (i.e., the edges), the natural approach to finding a max degree node is to sample. A prominent approach to sampling graphs is "star sampling", wherein a sample consists of a node u and its immediate neighborhood, denoted $\mathcal{N}(u)$. While this approach is reasonable for some graphs, it is suboptimal in (many) others, precisely because it ignores *local information*, i.e., the degrees of the neighbors of the current node. Exploiting this local information is the focus of random walk graph search algorithms, which repeatedly select, often at random and with bias based on neighbor degrees, a neighbor of the current node.

The main idea in this paper is to combine these two fundamental approaches (random sampling and random walks) into a single algorithm, which we term the *self-avoiding walk-jump (SAWJ)* algorithm. SAWJ exploits local information by repeatedly following maximum degree neighbors, as in random *walk*, or *jumps* to a randomly selected node, as in sampling, when the local information fails to provide a path towards high degree nodes. SAWJ is particularly effective on so-called *assortative* graphs, where $\alpha > 0$ for $\alpha \in [-1, +1]$ is defined as the correlation of the degrees of the endpoints of a randomly selected edge.

B. Related work

The literature on random walks on graphs for estimating graph properties is too large to cover; we present a few highlights that were particularly influential on our thinking. Avin [2] was one of the first to propose random walks biased towards unvisited high degree nodes. Lee [3] proposed a self avoiding random walk with edge re-weighting. The Albatross Sampling (AS) algorithm introduced by Jin [4] is a Metropolis Hastings random walk with random vertex sampling. The Frontier Sampling (FS) algorithm introduced by Ribeiro [5] employs multiple walkers that visit nodes with the same frequency as a simple random walk. Yet collectively the walkers achieve a lower mean-squared error in estimating graph properties than a simple random walk. We have implemented AS and FS and compare them against SAWJ in §V as we are unaware of any existing algorithms specifically designed for finding a max degree node in a large graph.

This research has been supported by the National Science Foundation BIGDATA program under award #IIS-1250786. The contact author is Steven Weber, sweber@coe.drexel.edu.

C. Contributions

Our key contribution is the proposed SAWJ algorithm. First, we provide two mathematical models to approximate the expected search time of a simplified version of the algorithm using the theory of absorbing Markov chains. Second, we evaluate the performance of SAWJ relative to star sampling, Albatross Sampling (AS) [4], and Frontier Sampling (FS) [5] algorithms on a class of graphs with tunable assortativity, called assortative Erdős-Rényi (AER) random graphs.

The paper is organized as follows. The model is introduced in §II, AER graphs and the Markov chain models of algorithm performance are given in §III and §IV respectively, the SAWJ algorithm and its performance are presented in §V, and a brief conclusion is given in §VI.

II. Model

A. Notation conventions

We write $[n]$ to denote $\{1,\dots,n\}$, for $n \in \mathbb{N}$. Sets are denoted in a calligraphic font, e.g., $\mathcal{D},\mathcal{E},\mathcal{V}$, etc., and often sets are partitioned into subsets with either single or double index, e.g., $\mathcal{E}_s,\mathcal{E}_{s,t}$. Collections of subsets are denoted in a script font, e.g., $\mathscr{E} = (\mathcal{E}_s, s \in \mathcal{D})$ denotes the collection of subsets $\mathscr{E}$ with elements (subsets) $\mathcal{E}_s$ indexed by s with index set $\mathcal{D}$. Quantities with a single (double) index are treated as a vector (matrix) and are denoted by a bold lower- (upper-) case roman symbol, e.g., $\mathbf{m} = (m_s, s \in \mathcal{D})$, and $\mathbf{M} = (M_{s,t}, (s,t) \in \mathcal{D}^2)$.

Random variables (RVs) are denoted in capital letter sans-serif font, $\mathsf{U},\mathsf{X},\mathsf{Y},\mathsf{Z}$, etc., with expectation ($\mathbb{E}[\cdot]$), probability ($\mathbb{P}(\cdot)$), covariance ($\mathrm{Cov}(\cdot,\cdot)$), correlation ($\mathrm{Corr}(\cdot,\cdot)$), variance ($\mathrm{Var}(\cdot)$), and standard deviation ($\mathrm{Std}(\cdot)$). We write $\mathsf{U} \sim \mathrm{Uni}(\mathcal{V})$ to denote a discrete uniform RV drawn from $\mathcal{V}$, $\mathsf{X} \sim \mathrm{Bin}(n,\rho)$ to denote a binomial RV with parameters (n,ρ), $\mathsf{Z} \sim \mathcal{N}(\mu,\sigma^2)$ to denote a normal RV with mean μ and variance σ^2, and we let $Q(\cdot)$ denote the standard normal CDF, i.e., $Q(z) \equiv \mathbb{P}(\mathsf{Z} \le z)$ for $\mathsf{Z} \sim \mathcal{N}(0,1)$. We use acronyms PDF/PMF for probability density/mass function, CDF for cumulative distribution function, and IID for independent and identically distributed. The notation $\mathsf{X} \approx \mathsf{Y}$ indicates that the RVs (X,Y) have approximately equal distribution.

B. Graph, neighborhood, degree, and degree distribution

We consider an undirected simple graph $G = (\mathcal{V},\mathcal{E})$ with $|\mathcal{V}| = n$ nodes and $|\mathcal{E}| = m$ edges. The set of edges is a collection of unordered pairs from $\mathcal{V}$, with $uv \equiv \{u,v\} \in E$ denoting an edge connecting nodes u and v. The *neighborhood* of node u is denoted $\mathcal{N}(u) = \mathcal{N}_G(u) \equiv \{v \in \mathcal{V} : uv \in \mathcal{E}\}$, and the *degree* of node u is the number of neighbors: $d(u) = d_G(u) \equiv |\mathcal{N}(u)|$. Let $\mathcal{D} \equiv d(\mathcal{V})$, for $d(\mathcal{V})$ the image of $d : \mathcal{V} \to \mathbb{Z}_+$, denote the *set of degrees* found in the graph. Let $\delta = \delta_G \equiv \max_{v\in\mathcal{V}} d(v) = \max \mathcal{D}$ be the *maximum degree*, and $\mathcal{V}_\delta \equiv \mathrm{argmax}_{v\in\mathcal{V}}\, d(v)$ be the *maximum degree nodes*. Partition the nodes $\mathcal{V}$ by degree into a collection of subsets $\mathscr{V} \equiv (\mathcal{V}_s, s \in \mathcal{D})$, with $\mathcal{V}_s \equiv \{v \in \mathcal{V} : d(v) = s\}$ the nodes of degree s. The *degree distribution* is denoted by $\mathbf{p} \equiv (p_s, s \in \mathcal{D})$, with $p_s \equiv |\mathcal{V}_s|/n$ the fraction of nodes of degree s. The *average degree*, over $v \in V$, is $\mu \equiv \sum_{s\in\mathcal{D}} s p_s = 2m/n$.

TABLE I
Notation

Symbol	Meaning
$G = (\mathcal{V},\mathcal{E})$	undirected graph
$\mathcal{V}, \lvert\mathcal{V}\rvert = n$	set (number) of nodes
$\mathcal{E}, \lvert\mathcal{E}\rvert = m$	set (number) of edges
$\mathcal{N}(u)$	neighbors of u
$d(u) = \lvert\mathcal{N}(u)\rvert$	degree of u
$\mathcal{D}$	set of degrees in G
δ	maximum degree
$\mathcal{V}_\delta$	maximum degree nodes
$\mathscr{V} = (\mathcal{V}_s, s \in \mathcal{D})$	partition of $\mathcal{V}$ by degree s
$\mathbf{p} = (p_s, s \in \mathcal{D})$	degree disbn., fraction of nodes of degree s
μ	average (over $\mathcal{V}$) degree
$\bar{\mathcal{V}}, \bar{n}$	set (number) of NI nodes
$\bar{\mathcal{D}}$	set of NI degrees in G
$\bar{\mathbf{p}} = (\bar{p}_s, s \in \bar{\mathcal{D}})$	NI degree disbn., fraction of nodes of degree s
$\bar{\mu}$	average (over $\bar{\mathcal{V}}$) NI degree
$\mathcal{E}_{s,t}, \mathcal{E}_{s,s}$	set of edges with degrees $\{s,t\}$, $\{s,s\}$
$\mathbf{F} = (F_{s,t})$	edge joint degree disbn.
$\bar{\mathcal{E}}_s$	edges with either/both nodes of degree s
$\mathbf{r} = (r_s, s \in \bar{\mathcal{D}})$	marginal disbn. for $\mathbf{F}$
ν, σ^2	mean and variance for disbn. $\mathbf{r}$
$\tilde{\mathbf{F}} = (\tilde{F}_{s,t})$	conditional disbn. of edge degree
$d_{\max}(u)$	maximum neighbor degree of node u
$\mathcal{V}_{s,t}$	nodes of degree s w. max. neighbor degree t
$\mathbf{H} = (H_{s,t})$	conditional maximum neighbor degree disbn.
$\alpha = \mathrm{Corr}(\mathsf{X},\mathsf{Y})$	assortativity: degree correlation on random edge
$r_1(\mathcal{E}), r_2(\mathcal{E})$	possible graph edges after primitive rewiring
Δ_1, Δ_2	changes in graph assortativity after rewiring
$\mathsf{G}_0 = (\mathcal{V},\mathsf{E}_0)$	Erdős-Rényi (ER) random graph w. params (n,ρ)
$\mathsf{G} = (\mathcal{V},\mathsf{E})$	assortative ER random graph w. params (n,ρ,α)
$\xi(n,\rho), \tilde{\xi}(n,\rho)$	approx. expected maximum degree $\mathbb{E}[\delta_{\mathsf{G}_0}]$
$\tilde{\mathbf{H}} = (\tilde{H}_{s,t})$	approximate $\mathbf{H}$ disbn. for AER graph
$\mathcal{N}_{\max}(u)$	set of maximum degree neighbors of u
$\mathsf{T}, \mathsf{D}, \mathsf{X}, \mathsf{A}, \mathsf{T}, \mathbf{q}, \mathbf{P}$	absorption time and Markov chain parameters
$\mathbf{h} = (h_s, s \in \mathcal{D})$	probability a degree s node is a local max

The set of isolated notes are those with zero degree $\mathcal{V}_0$. Non-isolated (henceforth NI) nodes are captured as follows. Let $\bar{\mathcal{V}} \equiv \mathcal{V} \setminus \mathcal{V}_0 = \{v \in \mathcal{V} : d(v) > 0\}$ be the *NI nodes*, $\bar{n} \equiv |\bar{\mathcal{V}}|$ the number of such nodes, and $\bar{\mathcal{D}} \equiv \mathcal{D} \setminus \{0\}$ the set of degrees found among NI nodes. Define the *NI degree distribution* $\bar{\mathbf{p}} \equiv (\bar{p}_s, s \in \bar{\mathcal{D}})$ with entries $\bar{p}_s \equiv |\mathcal{V}_s|/\bar{n}$. The *average NI degree*, over $v \in \bar{\mathcal{V}}$, is $\bar{\mu} \equiv \sum_{s\in\bar{\mathcal{D}}} s\bar{p}_s = 2m/\bar{n}$.

C. Joint and marginal degree distribution

Partition the edges $\mathcal{E}$ into the collection of subsets $\mathscr{E} = (\mathcal{E}_{s,t}, (s,t) \in \bar{\mathcal{D}}^2)$ with elements

$$\begin{aligned} \mathcal{E}_{s,t} &\equiv \{uv \in \mathcal{E} : \{d(u),d(v)\} = \{s,t\}\},\ s \neq t \\ \mathcal{E}_{s,s} &\equiv \{uv \in \mathcal{E} : d(u) = d(v) = s\} \end{aligned} \tag{1}$$

Here $\mathcal{E}_{s,t}$ are edges with degrees $\{s,t\}$, and $\mathcal{E}_{s,s}$ are edges with both endpoints of degree s. Define the *symmetric* $|\bar{\mathcal{D}}| \times |\bar{\mathcal{D}}|$ *joint degree distribution* matrix $\mathbf{F}$ with entries $F_{s,t}$:

$$F_{s,t} \equiv \begin{cases} \frac{|\mathcal{E}_{s,t}|}{2m}, & s \neq t \\ \frac{|\mathcal{E}_{s,s}|}{m}, & s = t \end{cases} \tag{2}$$

and observe $\sum_{s,t} F_{s,t} = 1$. We also have use for the collection $\bar{\mathscr{E}} \equiv (\bar{\mathcal{E}}_s, s \in \bar{\mathcal{D}})$ with elements $\bar{\mathcal{E}}_s \equiv \{uv \in \mathcal{E} : d(u) =$

s or $d(v) = s\}$ denoting the set of edges with *one or both* endpoints of degree s.

The marginal degree distribution $\mathbf{r} \equiv (r_s, s \in \bar{\mathcal{D}})$ corresponding to the joint degree distribution $\mathbf{F}$ is comprised of the row sums from $\mathbf{F}$:

$$r_s \equiv \sum_{t \in \bar{\mathcal{D}}} F_{s,t} = \frac{1}{2} \frac{|\bar{\mathcal{E}}_s| - |\mathcal{E}_{s,s}|}{m} + \frac{|\mathcal{E}_{s,s}|}{m}, \ s \in \bar{\mathcal{D}}. \quad (3)$$

Here r_s is the probability that a randomly chosen vertex of an edge selected uniformly at random from E has degree s. It is a standard result that

$$r_s = \frac{s\bar{p}_s}{\sum_{t \in \bar{\mathcal{D}}} t\bar{p}_t}, \ s \in \bar{\mathcal{D}}, \quad (4)$$

meaning that selecting a random vertex by first selecting a random edge biases the selection towards vertices of higher degree. To see (4), observe the following two quantities are equal: i) (4) equals $\frac{s|\mathcal{V}_s|}{2m}$, and ii) (3) equals $\frac{|\bar{\mathcal{E}}_s| + |\mathcal{E}_{ss}|}{2m}$, and both expressions equal the fraction of the $2m$ edge "stubs" connected to a node of degree s. The corresponding mean and variance of $\mathbf{r}$ are given by the usual formulae:

$$\nu \equiv \sum_{s \in \bar{\mathcal{D}}} s r_s, \ \sigma^2 \equiv \sum_{s \in \bar{\mathcal{D}}} (s - \nu)^2 r_s. \quad (5)$$

Finally, define the conditional degree distribution $\tilde{\mathbf{F}}$ obtained by normalizing each row of $\mathbf{F}$ so that it sums to one (so $\tilde{\mathbf{F}}$ is a stochastic matrix). In particular, $\tilde{F}_{s,t} \equiv F_{s,t}/r_s$. Write $\tilde{\mathbf{F}}_s = (\tilde{F}_{s,t}, t \in \bar{\mathcal{D}})$ for row s of $\tilde{\mathbf{F}}$. Note $\tilde{\mathbf{F}}_s$ is the conditional distribution of the other endpoint of a randomly selected edge for which one endpoint has degree s.

D. Max neighbor degree distribution and degree assortativity

For any NI node $u \in \bar{\mathcal{V}}$, define $d_{\max}(u) \equiv \max_{v \in \mathcal{N}(u)} d(v)$ as the maximum degree over the neighbors of u, hereafter called the *maximum neighbor degree*. Define the set of nodes with degree s and maximum neighbor degree t as

$$\mathcal{V}_{s,t} \equiv \{u \in \mathcal{V} : d(u) = s, d_{\max}(u) = t\}, \ (s,t) \in \bar{\mathcal{D}}^2. \quad (6)$$

Define the $|\bar{\mathcal{D}}| \times |\bar{\mathcal{D}}|$ conditional distribution matrix $\mathbf{H}$ with entries

$$H_{s,t} \equiv \frac{|\mathcal{V}_{s,t}|}{|\mathcal{V}_s|}, \ (s,t) \in \bar{\mathcal{D}}^2, \quad (7)$$

where $H_{s,t}$ is the probability that a randomly selected NI node of degree s has maximum neighbor degree t. Observe $\sum_{t \in \bar{\mathcal{D}}} H_{s,t} = 1$ for each $s \in \bar{\mathcal{D}}$. Let $\mathsf{Y} = d(\mathsf{U})$ be the random degree of a randomly selected NI node $\mathsf{U} \sim \mathrm{Uni}(\bar{\mathcal{V}})$, and let $\mathsf{M} = d_{\max}(\mathsf{U})$ be the corresponding maximum neighbor degree. Then $\mathbb{P}(\mathsf{M} = t | \mathsf{Y} = s) = H_{s,t}$.

Fix a graph $G = (\mathcal{V}, \mathcal{E})$, and let (X, Y) be the degrees of a randomly selected edge.

Definition 1. *Graph* assortativity[1] *is the correlation of the degrees of a randomly selected edge,* $\alpha \equiv \mathrm{Corr}(\mathsf{X}, \mathsf{Y})$.

[1]Note, the degree assortativity of a graph was first defined by Newman [6] in terms of the *joint excess degree distribution*; it is straightforward to see that our definition in terms of the *joint degree distribution* $\mathbf{F}$ is equivalent.

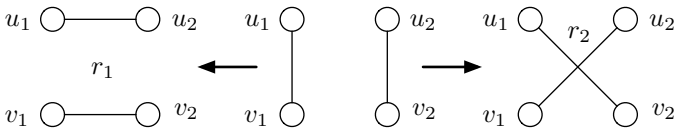

Fig. 1. Illustration of the primitive rewiring operation in Def. 2. The two edges u_1v_1 and u_2v_2 (middle) are replaced either with u_1u_2 and v_1v_2 (rewiring r_1, left) or with u_1v_2 and u_2v_1 (rewiring r_2, right), provided neither of the new edges are already present. Rewiring leaves all degrees unchanged.

Since X, Y are identically distributed:

$$\alpha \equiv \frac{\mathbb{E}[\mathsf{XY}] - \mathbb{E}[\mathsf{X}]^2}{\mathrm{Var}(\mathsf{X})} = \frac{1}{\sigma^2} \left(\sum_{(s,t) \in \bar{\mathcal{D}}^2} st F_{s,t} - \nu^2 \right). \quad (8)$$

The following lemma offers insight into the above sum.

Lemma 1. *Let* $G = (\mathcal{V}, \mathcal{E})$ *be a graph with* $|\mathcal{E}| = m$*, NI degrees* $\bar{\mathcal{D}}$*, and joint degree distribution* $\mathbf{F}$*. Then*

$$\sum_{(s,t) \in \bar{\mathcal{D}}^2} st F_{s,t} = \mathbb{E}[\mathsf{XY}] = \frac{1}{m} \sum_{uv \in \mathcal{E}} d(u) d(v). \quad (9)$$

Here (X, Y) are the degrees of a randomly selected edge, and the right side is the product of the degrees of the endpoints of an edge, averaged over all edges. The lemma follows by substituting the appropriate definitions.

E. Degree preserving rewiring

We define the following primitive rewiring operation on a graph $G = (\mathcal{V}, \mathcal{E})$, also used in [7] and our prior work [1].

Definition 2. *The primitive rewiring operation replaces any two edges* u_1v_1 *and* u_2v_2 *from* $\mathcal{E}$ *with either* i) u_1u_2 *and* v_1v_2 *(provided neither of these edges are already in* $\mathcal{E}$*) or* ii) u_1v_2 *and* u_2v_1 *(provided neither of these edges are already in* $\mathcal{E}$*):*

$$\begin{aligned} \mathcal{E} &\to r_1(\mathcal{E}) \equiv \mathcal{E} \setminus \{u_1v_1, u_2v_2\} \cup \{u_1u_2, v_1v_2\} \\ \mathcal{E} &\to r_2(\mathcal{E}) \equiv \mathcal{E} \setminus \{u_1v_1, u_2v_2\} \cup \{u_1v_2, u_2v_1\} \end{aligned} \quad (10)$$

These operations are illustrated in Fig. 1. The operation is degree preserving, i.e., the degree of each node is unchanged by rewiring. Let $G = (\mathcal{V}, \mathcal{E})$ be the graph before the operation, fix edges u_1v_1 and u_2v_2, and let $G_1 = G(\mathcal{V}, r_1(\mathcal{E}))$ and $G_2 = G(\mathcal{V}, r_2(\mathcal{E}))$ denote the two possible graphs after the operation. Let $\alpha(G), \alpha(G_1), \alpha(G_2)$ denote the corresponding assortativities. The following lemma computes the impact of the primitive rewiring operation on the assortativity.

Lemma 2. *Let* $G = (\mathcal{V}, \mathcal{E})$ *be a graph with* $|\mathcal{E}| = m$*, and fix two edges* u_1v_1 *and* u_2v_2*. Then the changes in assortativity under the two possible rewirings are* $\Delta_i \equiv \alpha(G_i) - \alpha(G)$*, for* $i \in \{1, 2\}$*, where*

$$\begin{aligned} \Delta_1 &= \frac{1}{m} \left(d(u_1)d(u_2) + d(v_1)d(v_2) - C \right) \\ \Delta_2 &= \frac{1}{m} \left(d(u_1)d(v_2) + d(u_2)d(v_1) - C \right) \end{aligned} \quad (11)$$

and $C \equiv d(u_1)d(v_1) + d(u_2)d(v_2)$.

Proof: The impact of rewiring on $\alpha(G_i)$ in (8) is restricted to the summation, by virtue of the fact that the rewiring is degree preserving, and therefore ν, σ^2 are unaffected. By Lem. 1, the only edges that change under the rewiring are the two removed edges and the two added edges. ∎

The term $d(u_1)d(v_1)+d(u_2)d(v_2)$ reflects the two removed edges, while $d(u_1)d(u_2) + d(v_1)d(v_2)$ and $d(u_1)d(v_2) + d(u_2)d(v_1)$ reflect the two possible pairs of new edges.

III. Assortative Erdős Rényi (AER) Graphs

A (non-assortative) Erdős-Rényi (ER) random graph $\mathsf{G}_0 = (\mathcal{V}, \mathsf{E}_0)$ has parameters (n, ρ), with $n \in \mathbb{N}$ and $\rho \in (0, 1)$. The node set is $\mathcal{V} = [n]$, and the *random* edge set E_0 is formed by adding each of the $\binom{n}{2}$ possible edges independently with probability ρ. Let $(\mathsf{X}_u, u \in \mathcal{V})$ be the random degrees in G_0; by construction it is clear that $\mathsf{X}_u \sim \mathrm{Bin}(n-1, \rho)$ for each $u \in \mathcal{V}$. Conditioned on $uv \in \mathsf{E}_0$, their random degrees (X, Y) each have distribution $1 + \mathrm{Bin}(n-2, \rho)$, where each binomial RV is formed from a separate set of (independent) $n-2$ Bernoulli RVs, and as such (X, Y) are (conditionally) independent. Thus the assortativity α_0 for an ER graph $\mathsf{G}_0 = (\mathcal{V}, \mathsf{E}_0)$ is 0.

A. Definition of an AER graph

Definition 3. *An assortative Erdős-Rényi (AER) random graph $\mathsf{G} = (\mathcal{V}, \mathsf{E})$ has parameters (n, ρ, α), with (n, ρ) as defined for an ER graph above, and edge degree correlation parameter $\alpha \in (-1, +1)$. It has nodes $\mathcal{V} = [n]$, and the random edges E have degrees (X, Y) with mean correlation $\mathrm{Corr}(\mathsf{X}, \mathsf{Y}) = \alpha$.*

The AER graph $\mathsf{G} = \mathsf{G}_k$ is obtained from an initial non-assortative ER graph $\mathsf{G}_0 = (\mathcal{V}, \mathsf{E}_0)$ by applying the primitive rewiring operation in Def. 2 k times in succession, each time selecting a pair of edges to rewire, yielding a sequence of graphs $\mathsf{G}_1, \ldots, \mathsf{G}_k$ with $\mathsf{G}_j = (\mathcal{V}, \mathsf{E}_j)$ and either $\mathsf{E}_j = r_1(\mathsf{E}_{j-1})$ or $\mathsf{E}_j = r_2(\mathsf{E}_{j-1})$ for $j \in [k]$.

The assortativity parameter $\alpha = \alpha_k$ for the final AER graph G is obtained by repeatedly applying Lem. 2, yielding

$$\alpha_k = \alpha_0 + \sum_{j=1}^{k} \Delta_{j,i(j)}, \tag{12}$$

where, as discussed above, $\alpha_0 = 0$, and $\Delta_{j,i(j)}$ is the change in correlation induced by rewiring the edge set from E_{j-1} to E_j, where $i(j) \in \{1, 2\}$ indicates which of the two rewirings in Def. 2 is selected in stage j.[2]

B. Joint degree distribution of an AER graph edge

Lemma 3. *If (X, Y) are the random degrees for a randomly selected edge from the AER graph $\mathsf{G} = (V, \mathsf{E})$ with parameters (n, ρ, α), then X and Y each have marginal binomial distribution $1+\mathrm{Bin}(n-2, \rho)$, and correlation α. The mean vector and covariance matrix for (X, Y) are given by $(\boldsymbol{\kappa}_n, \mathbf{C}_n)$, where:*

$$\begin{aligned} \boldsymbol{\kappa}_n &= (1 + (n-2)\rho) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ \mathbf{C}_n &= (n-2)\rho(1-\rho) \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \end{aligned} \tag{13}$$

The multivariate de Moivre Laplace theorem [8], gives conditions under which the multivariate binomial distribution converges to the multivariate normal.

Theorem 1 (Multivariate de Moivre Laplace [8])**.** *Let $\{\mathbf{X}_n\}$ be a sequence of k-dimensional binomial random vectors where $\mathbf{X}_n$ has mean $\boldsymbol{\kappa}_n$ and covariance matrix $\mathbf{C}_n$. Then $\mathbf{C}_n^{-1/2}(\mathbf{X}_n - \boldsymbol{\kappa}_n)$ converges in distribution to the standard k-variate normal random vector iff $\|\mathbf{C}_n^{-1}\| \to 0$ as $n \to \infty$.*

Observe that for $\mathbf{C}_n$ in Lem. 3 we have

$$\mathbf{C}_n^{-1} = \frac{1}{(1 + (n-2)\rho)(1-\alpha^2)} \begin{bmatrix} 1 & -\alpha \\ -\alpha & 1 \end{bmatrix} \tag{14}$$

and consequently $\|\mathbf{C}_n^{-1}\| \to 0$ as $n \to \infty$, for any matrix norm, as required by the theorem. As such, standardizing the binomial RVs as in the theorem yields, for large n, a pair of RVs that are approximately standard normal. Define the pair of normal RVs $(\tilde{\mathsf{X}}, \tilde{\mathsf{Y}}) \sim \mathcal{N}(\boldsymbol{\kappa}_n, \mathbf{C}_n)$, for $(\boldsymbol{\kappa}_n, \mathbf{C}_n)$ in Lem. 3. Thm. 1 gives the large-n approximation $(\mathsf{X}, \mathsf{Y}) \approx (\tilde{\mathsf{X}}, \tilde{\mathsf{Y}})$.

We next require a standard result for the bivariate normal distribution giving the distribution of one of the component random variables conditioned on the value of the other.

Lemma 4 ([9] §2.3)**.** *Let $(\mathsf{Z}_1, \mathsf{Z}_2)$ be bivariate normal with:*

$$(\mathsf{Z}_1, \mathsf{Z}_2) \sim \mathcal{N}\left(\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \gamma\sigma_1\sigma_2 \\ \gamma\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right). \tag{15}$$

Conditioned on $\mathsf{Z}_2 = z_2$, Z_1 is normally distributed, i.e., $\mathsf{Z}_1|\mathsf{Z}_2 = z_2 \sim \mathcal{N}(\eta_1(z_2), \sigma_1^2(z_2))$, where

$$\begin{aligned} \eta_1(z_2) &\equiv \mathbb{E}[\mathsf{Z}_1|\mathsf{Z}_2 = z_2] = \eta_1 + \gamma\frac{\sigma_1}{\sigma_2}(z_2 - \eta_2) \\ \sigma_1^2(z_2) &\equiv \mathrm{Var}(\mathsf{Z}_1|\mathsf{Z}_2 = z_2) = \sigma_1^2(1-\gamma^2). \end{aligned} \tag{16}$$

Thm. 1 and Lem. 4 yield the following corollary.

Corollary 1. *Let (X, Y) be the (binomial) random degrees for a randomly selected edge from the AER random graph $\mathsf{G} = (V, \mathsf{E})$ with parameters (n, ρ, α), and let $(\tilde{\mathsf{X}}, \tilde{\mathsf{Y}})$ be the normal approximation to (X, Y) from Thm. 1. Then, conditioned on $\mathsf{Y} = s$, Lem. 4 gives (with $\bar{\alpha} = 1 - \alpha$, $\bar{\rho} = 1 - \rho$):*

$$\begin{aligned} \mathbb{E}[\mathsf{X}|\mathsf{Y} = s] \approx \mathbb{E}[\tilde{\mathsf{X}}|\tilde{\mathsf{Y}} = s] &= \alpha s + \bar{\alpha}(1 + (n-2)\rho) \\ \mathrm{Var}(\mathsf{X}|\mathsf{Y} = s) \approx \mathrm{Var}(\tilde{\mathsf{X}}|\tilde{\mathsf{Y}} = s) &= (n-2)\rho\bar{\rho}(1-\alpha^2) \end{aligned} \tag{17}$$

C. Maximum neighbor degree distribution

Our next goal is to derive an approximation for the maximum neighbor degree conditional distribution matrix $\mathbf{H}$ for an AER random graph $\mathsf{G} = (V, \mathsf{E})$ with parameters (n, ρ, α). Let $\mathsf{Y} = d(\mathsf{U})$ be the degree of an NI node U selected uniformly

[2] It is certainly not the case that a graph *exists* for every choice of (n, α), and it is an interesting and (to our knowledge) open question to identify the set $\mathcal{A}(n) \subseteq [-1, +1]$ such that a graph with $|\mathcal{V}| = n$ exists for each $\alpha \in \mathcal{A}(n)$. This question is not our focus, however, and we leave it for future work.

at random from $\bar{V}$, and let $\mathsf{M} = d_{\max}(\mathsf{U})$ be the maximum neighbor degree of U. Our key approximation is conditional independence among the degrees of neighboring nodes of a given node of degree s. That is, given U, we *approximate* that the neighbor degrees $(\mathsf{X}_v, v \in \mathcal{N}(\mathsf{U}))$ of U are conditionally independent of one another, while at the same time, the mean edge degree correlation is $\mathrm{Corr}(\mathsf{Y}, \mathsf{X}_v) = \alpha$ for $v \in \mathcal{N}(\mathsf{U})$.

Lemma 5. *Let $\mathsf{G} = (V, \mathsf{E})$ be an AER graph with parameters (n, ρ, α) and maximum degree δ, and recall $\mathbf{H} = (H_{s,t})$ defined in (7). Then $H_{s,t} \approx \tilde{H}_{s,t}$ with*

$$\tilde{H}_{s,t} \equiv \begin{cases} Q_{s,t} - Q_{s,t-1}, & t \in \bar{\mathcal{D}} \setminus \{\delta\} \\ 1 - Q_{s,\delta-1}, & t = \delta \end{cases} \tag{18}$$

where

$$Q_{s,t} \equiv Q\left(\frac{t - \alpha s - (1-\alpha)(1 + (n-2)\rho)}{\sqrt{(n-2)\rho(1-\rho)(1-\alpha^2)}} \right)^s, \tag{19}$$

and $Q(\cdot)$ is the CDF of the standard normal distribution.

Proof: Fix a random NI node $\mathsf{U} \sim \mathrm{Uni}(\bar{V})$ with degree $\mathsf{Y} = d(\mathsf{U})$. Conditioned on $\mathsf{Y} = s$, let $(\mathsf{X}_v, v \in \mathcal{N}(\mathsf{U}))$ be the neighbor degrees, with each $\mathsf{X}_v \sim 1 + \mathrm{Bin}(n-2, \rho)$, and let $\mathsf{M} = \max_{v \in \mathcal{N}(\mathsf{U})} \mathsf{X}_v$ be the maximum neighbor degree. The edge-average correlation $\mathrm{Corr}(\mathsf{X}_v, \mathsf{Y})$ is α; we approximate that *each* edge has this correlation. We further approximate by treating the RVs (X_v) as conditionally independent of one another. With this and the normal approximation from Cor. 1,

$$\begin{aligned} H_{s,t} &= \mathbb{P}(\mathsf{M} = t | \mathsf{Y} = s) \\ &= \mathbb{P}(\mathsf{M} \le t | \mathsf{Y} = s) - \mathbb{P}(\mathsf{M} \le t-1 | \mathsf{Y} = s) \\ &\approx \mathbb{P}(\mathsf{X} \le t | \mathsf{Y} = s)^s - \mathbb{P}(\mathsf{X} \le t-1 | \mathsf{Y} = s)^s \\ &\approx \mathbb{P}(\tilde{\mathsf{X}} \le t | \tilde{\mathsf{Y}} = s)^s - \mathbb{P}(\tilde{\mathsf{X}} \le t-1 | \tilde{\mathsf{Y}} = s)^s \end{aligned} \tag{20}$$

Now standardize the conditional normal RV $\tilde{\mathsf{X}} | \tilde{\mathsf{Y}} = s$:

$$\begin{aligned} Q_{s,t} &\equiv \mathbb{P}(\tilde{\mathsf{X}} \le t | \tilde{\mathsf{Y}} = s)^s \\ &= \mathbb{P}\left(\frac{\tilde{\mathsf{X}} - \mathbb{E}[\tilde{\mathsf{X}} | \tilde{\mathsf{Y}} = s]}{\mathrm{Std}(\tilde{\mathsf{X}} | \tilde{\mathsf{Y}} = s)} \le \frac{t - \mathbb{E}[\tilde{\mathsf{X}} | \tilde{\mathsf{Y}} = s]}{\mathrm{Std}(\tilde{\mathsf{X}} | \tilde{\mathsf{Y}} = s)} \right)^s \end{aligned} \tag{21}$$

to obtain (19). ∎

IV. A Markov chain model for max degree search

First we introduce the essential idea of SAWJ in the form of Alg. 1, Walk-jump (WJ). Then we recall the mean absorption time of a discrete-time Markov chain in §IV-A, create a Markov chain approximating Alg. 1 for an *arbitrary* graph in terms of its max neighbor degree distribution $\mathbf{H}$ in §IV-B, and create a second Markov chain approximating Alg. 1 for an AER graph in terms of the parameters (n, ρ, α) in §IV-C. The expected fraction of nodes that are local maxima is estimated for arbitrary and AER graphs in §IV-D, and results showing the accuracy of the approximations are in §IV-E.

Algorithm 1 Walk–jump: find any max-degree node $v \in \mathcal{V}_\delta$

```
require graph G = (V, E), max degree δ = max_{u∈V} d(u)
while max{d(u), d_max(u)} < δ (i.e., V_δ not found) do
  if d(u) < d_max(u) (i.e., have higher degree nghbr) then
    select random v from N_max(u) (i.e., walk)
  else (i.e., at local maximum)
    select random v from V (i.e., jump)
```

A. Mean absorbtion time for a discrete-time Markov chain

Let T denote the random absorption time of an absorbing discrete-time Markov chain (DTMC) $\mathsf{D} \equiv (\mathsf{D}(\tau), \tau \in \mathbb{N})$ with finite state space $\mathcal{X}$, where $\mathsf{D}(0)$ has distribution $\mathbf{q} \equiv (q_x, x \in \mathcal{X})$. Partition $\mathcal{X}$ into transient $\mathcal{T}$ and absorbing $\mathcal{A}$ states, and partition the transition probability matrix $\mathbf{P}$:

$$\mathbf{P} = \begin{array}{c} \\ \mathcal{T} \\ \mathcal{A} \end{array} \begin{array}{c} \begin{array}{cc} \mathcal{T} & \mathcal{A} \end{array} \\ \begin{bmatrix} \mathbf{P}_{\mathcal{T},\mathcal{T}} & \mathbf{P}_{\mathcal{T},\mathcal{A}} \\ \mathbf{0} & \mathbf{P}_{\mathcal{A},\mathcal{A}} \end{bmatrix} \end{array}. \tag{22}$$

Here $\mathbf{P}_{\mathcal{T},\mathcal{T}}, \mathbf{P}_{\mathcal{T},\mathcal{A}}, \mathbf{P}_{\mathcal{A},\mathcal{A}}$ are submatrices corresponding to transitions between transient states, from transient to absorbing states, and among absorbing states, respectively. The *fundamental matrix*, defined below, is the key to the mean absorption time, as shown in Thm. 2.

Definition 4. *The* fundamental matrix *for the DTMC D is $\mathbf{N} = (\mathbf{I} - \mathbf{P}_{\mathcal{T},\mathcal{T}})^{-1}$, for $\mathbf{I}$ the identity matrix. The* mean absorption time *starting from state x is*

$$\tau_x \equiv \mathbb{E}[\mathsf{T} | \mathsf{D}(0) = x] = \mathbb{E}[\min\{\tau \in \mathbb{N} : \mathsf{D}(\tau) \in \mathcal{A} | \mathsf{D}(0) = x)]. \tag{23}$$

Theorem 2. *([10] §3.3) The mean absorption time is $\mathbb{E}[\mathsf{T}] = \sum_{x \in \mathcal{X}} \tau_x q_x$, where $\tau_x = (\mathbf{N}\mathbf{1})_x$, and $\mathbf{1}$ is a vector of ones.*

B. A Markov chain model for an arbitrary graph

We use Thm. 2 to compute the mean time to find a node in $\mathcal{V}_\delta$ under Alg. 1. Fix a graph $G = (\mathcal{V}, \mathcal{E})$ and from it compute the set of degrees $\mathcal{D}$, the max degree δ, the maximum degree nodes $\mathcal{V}_\delta$, the degree distribution $\mathbf{p} = (p_s, s \in \mathcal{D})$, and the maximum neighbor degree distribution $\mathbf{H}$. Using the notation §IV-A, define the discrete-time Markov chain $\mathsf{D} = (\mathsf{D}(\tau), \tau \in \mathbb{N})$, where $\mathsf{D}(\tau)$ represents the (random) degree of the node occupied at time τ, the state space is $\mathcal{X} = \mathcal{D}$, the absorbing state is $\mathcal{A} = \{\delta\}$, the transient states are $\mathcal{T} = \mathcal{D} \setminus \{\delta\}$, and the initial distribution $\mathbf{q}$ is $\mathbf{p}$. The transition submatrix $\mathbf{P}_{\mathcal{T},\mathcal{T}}$ has entries $P_{s,t} \equiv \mathbb{P}(\mathsf{D}(\tau+1) = t | \mathsf{D}(\tau) = s)$, where, for $s \in \mathcal{T}$:

$$P_{s,t} = \begin{cases} p_t, & s = 0 \\ h_s p_t, & s > 0, t \le s \\ h_s p_t + H_{s,t}, & 0 < s < t < \delta \end{cases} \tag{24}$$

Here, $\mathbf{h} \equiv (h_s, s \in \mathcal{D})$ has $h_s \equiv \sum_{t \le s} H_{s,t}$ equal to the probability the max neighbor degree of a degree s node is s or lower, i.e., the probability such a node is a local maximum.

Theorem 3. *The approximate mean time to find a maximum degree node for the above graph G under Alg. 1, starting*

from a node selected uniformly at random from $\mathcal{V}$, is given by Thm. 2 for the absorbing Markov chain described above.

Proof: The theorem requires we justify the given $\mathbf{P}_{\mathcal{T},\mathcal{T}}$. Consider a transient state $s \in \mathcal{T}$. There are two types of state transitions: i) from s to some $t \in \mathcal{T}$ with $t > s$ when node u has a max neighbor degree $t = d_{\max}(u) > s$, or ii) from s to any value $t \in \mathcal{D}$ when the max neighbor degree of u is s or lower. Then, writing D, D' for $\mathsf{D}(\tau), \mathsf{D}(\tau+1)$, respectively, we condition on M the maximum neighbor degree:

$$\begin{aligned} P_{s,t} &= \mathbb{P}(\mathsf{D}' = t | \mathsf{M} \le s, \mathsf{D} = s)\mathbb{P}(\mathsf{M} \le s | \mathsf{D} = s) + \\ &\quad \mathbb{P}(\mathsf{D}' = t | \mathsf{M} > s, \mathsf{D} = s)\mathbb{P}(\mathsf{M} > s | \mathsf{D} = s). \quad (25) \end{aligned}$$

For any t, the first term simplifies to $p_t h_s$, since the event $\{\mathsf{M} \le s\}$ corresponds to node s being a local maximum node, which has probability $h_s = \mathbb{P}(\mathsf{M} \le s | \mathsf{D} = s)$, and results in the decision to jump to a random node, which will have degree t with probability p_t. For $t > s$ the second term simplifies to $H_{s,t} = \mathbb{P}(\mathsf{M} = t | \mathsf{D} = s)$, while for $t \le s$ the second term is zero, since the algorithm will never select a node of degree $t \le s$ when there is a neighbor of degree exceeding s. ∎

C. A Markov chain model for an AER random graph

Whereas Thm. 3 applies to an *arbitrary* graph, and requires knowledge of the max neighbor degree distribution $\mathbf{H}$ to compute the transition submatrix $\mathbf{P}_{\mathcal{T},\mathcal{T}}$, in this section we consider an AER random graph $\mathsf{G} = (\mathcal{V}, \mathsf{E})$ with parameters $(n, \rho, \alpha, \delta)$ if the maximum degree δ is known, or (n, ρ, α) given only an approximate maximum degree $\tilde{\xi} \equiv \tilde{\xi}(n, \rho) \approx \lfloor (n-1)\rho + \sqrt{2(n-1)(log\ n)\rho(1-\rho)} \rfloor$.

Using the notation in §IV-A, define the discrete-time Markov chain $\tilde{\mathsf{D}} = (\tilde{\mathsf{D}}(\tau), \tau \in \mathbb{N})$, where $\tilde{\mathsf{D}}(\tau)$ represents the (random) degree of the node occupied at time τ, the state space is $\mathcal{X} = \tilde{\mathcal{D}} = \{0, \ldots, \tilde{\xi}\}$, the absorbing state is $\mathcal{A} = \{\tilde{\xi}\}$, the transient states are $\mathcal{T} = \tilde{\mathcal{D}} \backslash \{\tilde{\xi}\}$, and the initial distribution $\mathbf{q}$ is $\tilde{\mathbf{p}} = (\tilde{p}_s, s \in \tilde{\mathcal{D}})$ with entries

$$\tilde{p}_s = \frac{\mathbb{P}(\mathsf{X} = s)}{\mathbb{P}(\mathsf{X} \le \tilde{\xi})},\ s \in \tilde{\mathcal{D}}, \quad (26)$$

for $\mathsf{X} \sim \mathrm{Bin}(n-1, \rho)$, i.e., $\tilde{p}_s$ is the probability a binomial RV equals s conditioned on the RV taking value at most $\tilde{\xi}$. The transition submatrix $\mathbf{P}_{\mathcal{T},\mathcal{T}}$ has entries $\tilde{P}_{s,t} \equiv \mathbb{P}(\tilde{\mathsf{D}}(\tau+1) = t | \tilde{\mathsf{D}}(\tau) = s)$, where, for $s \in \mathcal{T}$ (recall Lem. 5):

$$\tilde{P}_{s,t} = \begin{cases} \tilde{p}_t, & s = 0 \\ Q_{s,s}\tilde{p}_t, & s > 0, t \le s \\ Q_{s,s}\tilde{p}_t + Q_{s,t} - Q_{s,t-1}, & 0 < s < t < \tilde{\xi} \end{cases} \quad (27)$$

Theorem 4. *The approximate mean time to find a maximum degree node for the above AER random graph G under Alg. 1, starting from a node selected uniformly at random from $\mathcal{V}$, is given by Thm. 2 for the Markov chain described above.*

Proof: The proof of Thm. 4 follows from the proof of Thm. 3 by recognizing the approximations $h_s = \mathbb{P}(\mathsf{M} \le s | \mathsf{D} = s) \approx Q(s, s)$, $H_{s,t} \approx \tilde{H}_{s,t} \equiv Q_{s,t} - Q_{s,t-1}$ for $t < \tilde{\xi}$, and $H_{s,\tilde{\xi}} \approx 1 - Q_{s,\tilde{\xi}-1}$. ∎

D. Approximate fraction of local maxima nodes

Alg. 1 jumps at a local maximum node, i.e., a node with $d(u) \ge d_{\max}(u)$, and as such the absorption time is sensitive to the fraction of strict $f_{\mathrm{str}} \equiv |\mathcal{V}_{\mathrm{str}}|/\bar{n}$ and non-strict $f_{\mathrm{nst}} \equiv |\mathcal{V}_{\mathrm{nst}}|/\bar{n}$ local maxima, where $\mathcal{V}_{\mathrm{str}} \equiv \{u \in \bar{\mathcal{V}} : d(u) > d_{\max}(u)\}$, and $\mathcal{V}_{\mathrm{nst}} \equiv \{u \in \bar{\mathcal{V}} : d(u) \ge d_{\max}(u)\}$ are the corresponding sets of local maximum NI nodes. Isolated nodes are not considered maxima. We employ the same conditional independence approximation used in Thm. 3 and Thm. 4.

Proposition 1. *For a graph $G = (\mathcal{V}, \mathcal{E})$ with conditional maximum neighbor degree distribution $\mathbf{H}$:*

$$f_{\mathrm{str}} \approx \tilde{f}_{\mathrm{str}} = \sum_{s \in \bar{\mathcal{D}}} (h_s - H_{s,s})\bar{p}_s,\ f_{\mathrm{nst}} \approx \tilde{f}_{\mathrm{nst}} = \sum_{s \in \bar{\mathcal{D}}} h_s \bar{p}_s, \quad (28)$$

for $\mathbf{h} = (h_s, s \in \bar{\mathcal{D}})$ with $h_s = \sum_{t \le s} H_{s,t}$, and $\bar{\mathbf{p}}$ as in §II-B. For an AER graph $\mathsf{G} = (\mathcal{V}, \mathsf{E})$ with parameters (n, ρ, α):

$$f_{\mathrm{str}} \approx \hat{f}_{\mathrm{str}} = \sum_{s \in [\tilde{\xi}]} Q_{s,s-1}\tilde{p}_s,\ f_{\mathrm{nst}} \approx \hat{f}_{\mathrm{nst}} = \sum_{s \in [\tilde{\xi}]} Q_{s,s}\tilde{p}_s, \quad (29)$$

for $\tilde{\xi}$ in §IV-C, $Q_{s,t}$ in (19), and $\tilde{\mathbf{p}} = (\tilde{p}_s, s \in [\tilde{\xi}])$ in (26).

Proof: For the graph $G = (\mathcal{V}, \mathcal{E})$ with distribution $\mathbf{H}$, condition on the degree $\mathsf{Y} = d(\mathsf{U})$ of a node $\mathsf{U} \sim \mathrm{Uni}(\bar{\mathcal{V}})$, and let $\mathsf{M} = d_{\max}(\mathsf{U})$ be the maximum neighbor degree:

$$f_{\mathrm{str}} = \sum_{s \in \bar{\mathcal{D}}} \mathbb{P}(\mathsf{M} < s | \mathsf{Y} = s)\mathbb{P}(\mathsf{Y} = s) \approx \sum_{s \in \bar{\mathcal{D}}} (h_s - H_{s,s})\bar{p}_s. \quad (30)$$

For the AER graph $\mathsf{G} = (\mathcal{V}, \mathsf{E})$ with parameters (n, ρ, α):

$$f_{\mathrm{str}} = \sum_{s \in \bar{\mathcal{D}}} \mathbb{P}(\mathsf{M} < s | \mathsf{Y} = s)\mathbb{P}(\mathsf{Y} = s) \approx \sum_{s \in [\tilde{\xi}]} Q_{s,s-1}\tilde{p}_s, \quad (31)$$

following an identical derivation used in Lem. 5. The approximations for f_{nst} follow replacing $< s$ with $\le s$. ∎

E. Evaluation of the accuracy of the approximations

Fig. 2 shows numerical and simulation results illustrating the accuracy of the approximations used in obtaining the mean absorption time in Thm. 3, Thm. 4, and the fraction of local maxima nodes in Prop. 1 on AER graphs. The graphs have $n = 1000$ nodes and edge probability $\rho = 0.005$, yielding mean degree $\mu = 5$ and max degree of $\tilde{\xi} \approx 14$. We swept α over $[-0.8, +0.8]$, and created 10 AER graphs for each α.

The accuracy of Thm. 3 and Thm. 4 is shown in Fig. 2 (top). The red curve labeled "Walk-jump" is obtained by averaging the absorption time over 1000 trials of Alg. 1 on each graph and α value (for $10,000$ simulations per α). Curves "MC Model" and "AER MC Model" are found by computing the mean absorption time for the Markov chains in Thm. 3 (using matrix $\mathbf{H}$) and Thm. 4 (using parameters (n, ρ, α)), respectively. Note Thm. 3 is *very* accurate, and Thm. 4 is reasonably accurate.[3]

[3]The AER curve uses the *actual* δ_G in each graph instead of $\tilde{\xi}$, the estimate.

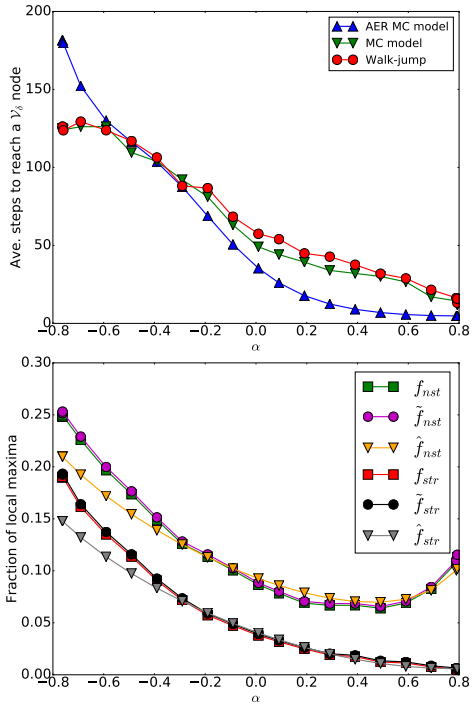

Fig. 2. Numerical and simulation results to measure the accuracy of the approximations used in the Markov models predicting the mean search time in Thm. 3 and Thm. 4 (top) and the mean fraction of local maxima nodes in Prop. 1 (bottom), vs. the assortativity α. See §IV-E for explanation.

The accuracy of Prop. 1 is shown in the six curves in Fig. 2 (bottom): three curves each for the fraction of strict and non-strict local maxima. For each set of three we show the actual fraction, the approximation for the graph using its maximum neighbor degree distribution $\mathbf{H}$ (denoted $\hat{f}$), and the AER approximation using (n, ρ, α) (denoted $\tilde{f}$). The results show both approximations are very accurate for a wide range of α.[4]

V. The Self-Avoiding Walk-Jump (SAWJ) Algorithm

The search time of Alg. 1 was approximated by the Markov chains in Thm. 3 and Thm. 4. We improve Alg. 1 in §V-A, describe several competing algorithms from the literature in §V-B, then study the performance (search time) of Alg. 2 against competing algorithms on both AER graphs (§V-C) and on real-world large graphs (§V-D).

A. Algorithm description

The self-avioding walk-jump (SAWJ) algorithm is given in Alg. 2. It requires a graph $G = (\mathcal{V}, \mathcal{E})$, the maximum degree δ_G, an algorithm bias parameter $b \in [0, 1]$, and assumes $\alpha > 0$. The key addition in Alg. 2 relative to Alg. 1 is the self avoidance of the walk. In Alg. 2, $\mathcal{N}_{max}(u)$ is the maximum degree neighbors of node u. If $\alpha < 0$ SAWJ, is modified to walk to minimum degree unvisited neighbors, under the supposition that $\mathcal{V}_\delta$ nodes in graphs with negative assortativity have *low* degree neighbors. The modified SAWJ algorithm jumps if it is at local minima and $d_{max}(u) < \delta$.

[4]The AER curve uses standard de Moivre Laplace integration corrections.

Algorithm 2 Self-avoiding walk–jump (SAWJ): find a $v \in \mathcal{V}_\delta$

require graph $G = (\mathcal{V}, \mathcal{E})$, max deg. δ_G, bias $b \in [0, 1]$
initialize $u \in \mathcal{V}$ *(starting node)*, $\mathcal{H} := \emptyset$ *(init. history)*
while $\max\{d(u), d_{\max}(u)\} < \delta_G$ *(not yet found $\mathcal{V}_\delta$)* **do**
  $\mathcal{H} := \mathcal{H} \cup \{u\}$ *(update history)*
  $d_{\max}(u) := \max_{v \in \mathcal{N}(u) \setminus \mathcal{H}} d(v)$ *(max neighbor degree)*
  $\mathcal{N}_{\max}(u) := \operatorname{argmax}_{v \in \mathcal{N}(u) \setminus \mathcal{H}} d(v)$ *(max deg. nghbrs)*
  if $\mathcal{N}_{\max}(u) \neq \emptyset$ **then** *(there are unvisited neighbors)*
    if $d_{\max}(u) \geq d(u)$ *(i.e., not local max)* **then**
      select random $v \in \mathcal{N}_{\max}(u)$ *(i.e., walk)*
    else *(i.e., at a local max)*
      w.p. b select random $v \in \mathcal{V} \setminus \mathcal{H}$ *(i.e., jump)*
      w.p. $1 - b$ select random $v \in \mathcal{N}_{\max}(u)$ *(i.e., walk)*
  else *(no unvisited neighbors)*
    select random $v \in \mathcal{V} \setminus \mathcal{H}$ *(i.e., jump)*

B. Competing algorithms

We compared the search algorithms proposed above against two "star" (neighborhood) sampling algorithms and two sampling (random walk) algorithms from the literature.

1) Star sampling with replacement (Star-R) [11] Repeatedly select a random node $u \in \mathcal{V}$ until $d(u) = \delta$ or $d_{\max}(u) = \delta$.
2) Star sampling without replacement (Star-NR) [11] A history $\mathcal{H}$ is maintained of visited nodes and their neighbors, and a random node $u \in \mathcal{V} \setminus \mathcal{H}$ is repeatedly selected until $d(u) = \delta$ or $d_{\max}(u) = \delta$.
3) Albatross sampling (AS) [4] repeatedly do the following: jump (with probability 0.2) to a random node or walk (w.p. 0.8) to a randomly selected neighbor of the current node.
4) Frontier sampling (FS) [5] employs m (we use $m = 5$) simultaneous random walkers on the graph; at each time one of the walkers is selected in proportion to their current degrees, and moves to a randomly selected neighbor.

To compare AS and FS with SAWJ, we terminate AS and FS if either are in the neighborhood of a $\mathcal{V}_\delta$ node.

C. Algorithm performance on AER graphs

The performance of the SAWJ and WJ algorithms from §V-A and of the competing Star-R, Star-NR, AS, and FS algorithms on AER graphs is shown in Fig. 3. The graphs have $n = 1000$ nodes and edge density $\rho = 0.005$ (top, with $(\mu, \tilde{\xi}) = (5, 14)$). The assortativity α is swept over $[-0.8, +0.8]$, and for each value of α we created 10 AER graphs. We then ran each of the 8 algorithms 1000 times on each graph, for $10,000$ trials per α. The figure shows the average search time to find a maximum degree node by each algorithm for a graph with a given α.

Several points bear mention. First, comparing SAWJ vs. WJ, observe the self-avoiding property and following $d_{\min}(u)$ can halve the search time for $\alpha \approx -1$, at $\alpha \approx 0$ WJ can outperform SAWJ, and for $\alpha > 0$ SAWJ slightly outperforms WJ. The $\alpha > 0$ case shows self-avoidance is beneficial, however the benefits may not be pronounced if both SAWJ and WJ revert to random sampling or follow an increasing degree gradient.

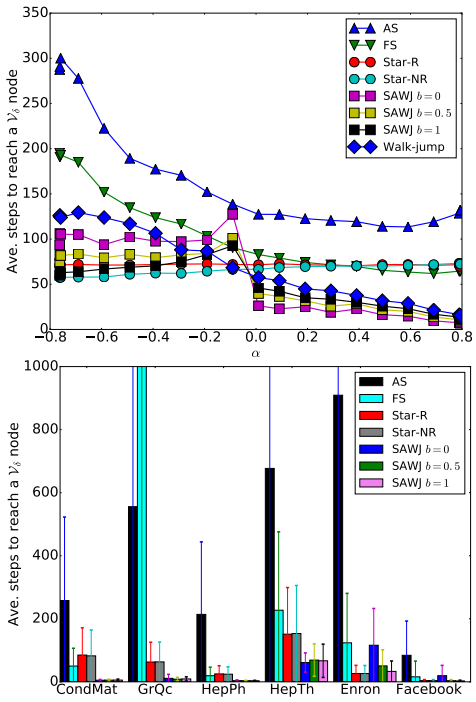

Fig. 3. The mean search time to find a max degree node in $\mathcal{V}_\delta$ for AER graphs with $n = 1000$ nodes and edge probabilities $\rho = 0.005$ ($\mu = 5$, $\xi = 14$) vs. the assortativity α (top). Mean number of steps to find a max degree node ($\mathcal{V}_\delta$) for the SNAP graphs (bottom).

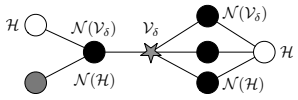

Fig. 4. Example showing star-sampling without replacement (Star-NR) may be inferior to star-sampling with replacement (Star-R). The two white nodes ($\mathcal{H}$) have been sampled; these two nodes and their four neighbors, the black nodes $\mathcal{N}(\mathcal{H})$, have been removed from the sampling pool. The maximum degree node is the star ($\mathcal{V}_\delta$); the black nodes are *also* the neighbors of $\mathcal{V}_\delta$, denoted $\mathcal{N}(\mathcal{V}_\delta)$. The probability of reaching $\mathcal{V}_\delta$ or its neighbors $\mathcal{N}(\mathcal{V}_\delta)$ on the next sample is $1/2$ without replacement and $5/8$ with replacement.

Second, observe SAWJ has a slight upward trend in α for $\alpha < 0$, a "spike" at $\alpha = 0$, and a downward trend in α for $\alpha > 0$. The improvement as $\alpha \uparrow 1$ ($\alpha \downarrow -1$) reflects the increased value of following maximum (minimum) degree neighbors for increasingly assortative (disassortative) graphs, respectively; the spike at $\alpha = 0$ suggests following a negative degree gradient is not optimal in slightly disassortative graphs. Third, we found the optimal choice of bias parameter b for the SAWJ algorithm depends upon (ρ, α), although $b = 0$ often performs well. Fourth, although Star-NR is optimal for low-density ($\rho = 0.005$) disassortative ($\alpha < 0$) graphs, SAWJ performs better on high-density or assortative graphs. The AS and FS algorithms are not competitive for most values of α.

A surprising observation is that in Fig. 3 we see star sampling without replacement (Star-NR) is not always superior to star sampling with replacement (Star-R). Fig. 4 gives an example to illustrate how this may occur.

D. Algorithm performance on real-world graphs

We tested the SAWJ algorithm on six real-world graphs from the Stanford Large Network Dataset Collection (SNAP) [12], see Table II. For all six graphs $|\mathcal{V}_\delta| = 1$. Averaging over $10,000$ trials per graph, SAWJ significantly outperforms star-sampling for *CondMat*, *GrQC*, *HepPH*, and *HepTh* where ($\alpha > 0$). In the *Facebook* graph ($\alpha \approx 0$) SAWJ $b = 1$ slightly outperforms star sampling. In the Enron graph ($\alpha < 0$) star sampling is superior to SAWJ. In all six graphs SAWJ outperformed AS and FS. In GrQc we terminated FS at 10 million steps if it had not found a $\mathcal{V}_\delta$ node; FS's mean time to find a $\mathcal{V}_\delta$ node in GrQc is approximately $3,500$ steps.

Graph	n	m	ρ	μ	δ	α
CondMat	23,133	93,497	0.00035	4.0	279	0.13
GrQc	5,242	14,496	0.00106	2.7	81	0.66
HepPh	12,008	118,521	0.00164	9.9	491	0.63
HepTh	9,877	25,998	0.00053	2.6	65	0.27
Enron	36,692	183,831	0.00027	5.0	1,383	-0.11
Facebook	4,039	88,234	0.01082	21.8	1,045	0.06

TABLE II
REAL-WORLD SNAP GRAPH [12] PROPERTIES.

VI. CONCLUSIONS

We have presented the SAWJ algorithm, designed to find maximum degree nodes in large assortative graphs. The algorithm was shown to perform well relative to Star-R, Star-NR, AS, and FS, both for a wide range of AER graphs and for six real-world large graphs from the Stanford SNAP dataset. We presented Markov models to approximate the required search time under a simplified version of SAWJ for both an arbitrary graph and an AER graph, and the corresponding absorption times are reasonably close to the simulated search times.

REFERENCES

[1] J. Stokes and S. Weber, "On random walks and random sampling to find max degree nodes in assortative Erdős Rényi graphs," in *IEEE Global Communications Conference*, Washington, D.C., December 2016.
[2] C. Avin and B. Krishnamachari, "The power of choice in random walks: an empirical study," *Computer Networks*, vol. 52, no. 1, pp. 44–60, 2008.
[3] C.-H. Lee, X. Xu, and D. Y. Eun, "Beyond random walk and Metropolis-Hastings samplers," *ACM Performance Evaluation Review*, vol. 40, no. 1, pp. 319–330, June 2012.
[4] L. Jin, Y. Chen, P. Hui, C. Ding, T. Wang, A. V. Vasilakos, B. Deng, and X. Li, "Albatross sampling: robust and effective hybrid vertex sampling for social graphs," in *ACM HotPlanet*, Bethesda, MD, 2011.
[5] B. Ribeiro and D. Towsley, "Estimating and sampling graphs with multi-dimensional random walks," in *ACM Internet Measurement Conference (IMC)*, Melbourne, Australia, November 2010, pp. 390–403.
[6] M. E. J. Newman, "Assortative mixing in networks," *Physics Review Letters*, vol. 89, no. 20, p. 208701, October 2002.
[7] L. Li, D. Alderson, J. C. Doyle, and W. Willinger, "Towards a theory of scale-free graphs: definition, properties, and implications," *Internet Mathematics*, vol. 2, no. 4, pp. 431–523, 2005.
[8] J. A. Veeh, "The multivariate Laplace-De Moivre theorem," *Journal of Multivariate Analysis*, vol. 18, no. 1, pp. 46–51, 1986.
[9] C. Chatfield and A. Collins, *Introduction to Multivariate Analysis*, 1st ed., ser. CRC Texts in Statistical Science. Chapman & Hall, 1980.
[10] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*. Springer, 1983.
[11] E. D. Kolaczyk, *Statistical analysis of network data : methods and models*. New York, London: Springer, 2009.
[12] J. Leskovec and A. Krevl, "Stanford Large Network Dataset Collection (SNAP)," http://snap.stanford.edu/data, June 2014.