A Proof of Theorem 1

Here, we give a full proof of Theorem 1. Recall that an equivalent construction of the PAGA graphs is obtained by collapsing the graph $G_{m}^{t}$ to $G_{m}^{t+1}$ leading to $t$ super nodes. Accordingly, based on $G_{m}^{t}$, we can directly generate $G_{m}^{t+1}$ by (temporarily) adding $m$ more nodes, each linking with a single edge, and then collapsing these $m$ nodes to the $(t+1)$th super node. Let $d_{i,t}(v)$ denote the degree of the super-node $v$ after $mt+i-1$ edges have been introduced, the probability that the $mt+i$th node (or equivalently its corresponding super-node $t+1$) connects to $v$ is

$$
Pr\{t+1,v\} = \begin{cases} 
\frac{d_{i,t}(v)+mf(v,t+1)-1}{(f(\cdot,t+1)+1)(mt+i)-1} & 1 \leq v \leq t \\
\frac{d_{i,t}(t+1)+f(t+1)-1+1}{(f(\cdot,t+1)+1)(mt+i)-1} & v = t+1.
\end{cases}
$$

(15)

Given these definitions, we begin with proving statement 1. With Eq. (9) and starting conditions

$$
N(k,1) = [k = 2m], \quad N(k,t) = N(k,t,1)
$$

we get a complete description of $N(k,t)$ as Eq. (8) of statement 1

$$
N(k,t+1) = N(k,t,m+1) + Pr(d_{i,m+1}(t+1) = k)
$$

Next we prove statement 2 which is straightforward from the definition. First, note that $k \in \{m, m+1, \ldots, 2m\}$, i.e.:

$$
Pr(d_{i,k+1}(t+1)k) = 0, \quad k < m \text{ or } k > 2m.
$$

(17)

The minimal value $d_{i,m+1}(t+1) = m$ is obtained when no one of the $k$ edges is a loop. In this case, $d_{i,i}(t+1) = i-1$ for all $i$, so:

$$
Pr(d_{i,m+1}(t+1) = m) = \prod_{i=1}^{m} \left(1 - \frac{i \cdot f(\cdot, t+1)}{(f(\cdot, t+1) + 1)(mt+i) - 1}\right)
$$

$$
= 1 + O_{m} \left(\frac{1}{tf(\cdot, t+1)}\right).
$$

From $\sum_{k=m}^{2m} Pr(d_{i,m+1} = k) = 1$ and $Pr(d_{i,m+1} = k) \geq 0$:

$$
Pr(d_{i,m+1}(t+1) = k) = O_{m} \left(\frac{1}{tf(\cdot, t+1)}\right), \quad m < d \leq 2m.
$$

(18)

Finally the claim follows from the definition of $[k = m]$.

Before proving statement 3, we first derive some properties of $c(k,t,i)$ and prove statement 4 en-route. Here, for brevity, we denote $c(k,t,i)$ as $c(k)$. Starting from $c(m)$ we study the step-wise change of $c(k)$ as:

$$
c(m) = \frac{B(m \cdot f(m,t,i), f(\cdot, t)+2)}{B(m \cdot f(m,t,i), f(\cdot, t+1))}
$$

$$
= \frac{\Gamma(f(\cdot, t)+2)}{\Gamma(f(\cdot, t+1))} \times \frac{\Gamma(m \cdot f(m,t,i)+f(\cdot, t)+1)}{\Gamma(m \cdot f(m,t,i)+f(\cdot, t)+2)}
$$

$$
= \frac{f(\cdot, t)+1}{m \cdot f(m,t,i)+f(\cdot, t)+1}
$$

(19)

For $k > m$, the ratio of $c(k,i)$ to $c(k-1,i)$ can be simplified as:

$$
\frac{c(k-1)}{c(k)} = \frac{B(k-1 + m(f(k-1,t,i)-1), f(\cdot, t)+2)}{B(k + m(f(k,t,i)-1)+f(\cdot, t+2))}
$$

$$
= \frac{\Gamma(k - 1 + m(f(k-1,t,i)-1))}{\Gamma(k + m(f(k,t,i)-1))} \times \frac{\Gamma(k + m(f(k,t,i)-1)+f(\cdot, t)+2)}{\Gamma(k-1 + m(f(k-1,t,i)-1)+f(\cdot, t)+2)}
$$

$$
= \frac{k + m(f(k,t)-1) + 1 + f(\cdot, t)}{k - 1 + m(f(k-1,t)-1)}
$$

(20)
Rewriting the \( N \) of (11) can be expressed in a recursive fashion as:

\[
\ln c(k,i) = C_0 + \ln \Gamma(k+C_1) - \ln \Gamma(k+C_1 + \mathcal{f}(\cdot, t) + 2)
\]

\[
C_0 = \ln \frac{\Gamma(\mathcal{f}(\cdot, t) + 2)}{B(m \cdot \mathcal{f}(m, t, i), \mathcal{f}(\cdot, t) + 1)}, C_1 = m(\mathcal{f}(k, t, i) - 1)
\]  \quad (21)

Exploiting the fact that asymptotically

\[
\Gamma(x+1) \sim \sqrt{2\pi x} \left(\frac{x}{e}\right)^x \ln \sqrt{2\pi} + x(\ln x - 1) + \frac{1}{2} \ln x,
\]

Eq. (21) becomes:

\[
C_0 + (k+C_1)(\ln(k+C_1) - 1) + O\left(\frac{1}{k}\right) - (k+C_1 + \mathcal{f}(\cdot, t) + 2)\left(\ln(k+C_1 + \mathcal{f}(\cdot, t) + 2) - 1\right)
\]

\[
= C_0 + (k+C_1)\left(\ln k + \frac{C_1}{k} - 1\right) + O\left(\frac{1}{k}\right) - (k+C_1 + \mathcal{f}(\cdot, t) + 2)\left(\ln k + \frac{C_1 + \mathcal{f}(\cdot, t) + 2}{k} - 1\right)
\]

\[
= C_0 - (2 + \mathcal{f}(\cdot, t)) \ln k + O\left(\frac{1}{k}\right)
\]

Rewriting the \( C_0 \) term, the full expression of asymptotic value of \( c(k) \) as \( k \) grows becomes

\[
c(k) = \frac{\Gamma(\mathcal{f}(\cdot, t) + 2)}{B(m \cdot \mathcal{f}(m, t, i), \mathcal{f}(\cdot, t) + 1)} k^{-2 \mathcal{f}(\cdot, t)} \left(1 + O\left(\frac{1}{k}\right)\right).
\]

To prove the statement 3, we show

\[
N(k, t, i + 1)
\]

\[
= \sum_{v=1}^{t} \Pr(d_{t,i+1}(v) = k)
\]

\[
= \sum_{v=1}^{t} \left(\Pr(d_{t,i+1}(v) = k, \gamma = v) + \Pr(d_{t,i+1}(v) = k, \gamma \neq v)\right)
\]

\[
= \sum_{v=1}^{t} \left(\Pr(d_{t,i}(v) = k-1, \gamma = v) + \Pr(d_{t,i}(v) = k, \gamma \neq v)\right)
\]

\[
= \sum_{v=1}^{t} \left(\Pr(d_{t,i}(v) = k-1) \frac{k-1+m(\mathcal{f}(v, t+1) - 1)}{\mathcal{f}(\cdot, t+1) + 1} + \Pr(d_{t,i}(v) = k) \left(1 - \frac{k+m(\mathcal{f}(v, t+1) - 1)}{\mathcal{f}(\cdot, t+1) + 1} \right)\right)
\]

\[
= N(k-1, t, i) \frac{k-1+m(\mathcal{f}(k-1, t+1, i) - 1)}{\mathcal{f}(\cdot, t+1) + 1} + N(k, t, i) \left(1 - \frac{k+m(\mathcal{f}(k, t+1, i) - 1)}{\mathcal{f}(\cdot, t+1) + 1} \right)
\]  \quad (22)

By proving statement 4, we complete our proof of Theorem 1. For the ease of proof, let us define \( \tilde{N}(k, t, i) \):

\[
\tilde{N}(k, t, i) = N(k, t, i) - c(k, t, i) \left(t + \frac{i}{m} - \frac{1}{m(1 + \mathcal{f}(\cdot, t))}\right)
\]

\[
(23)
\]

The lhs of (11) can be expressed in a recursive fashion as:

\[
\tilde{N}(k, t, i + 1) + \frac{i}{m} [k = m]
\]

\[
= \tilde{N}(k, t, i + 1) + \frac{1}{m} [k = m] + \frac{i-1}{m} [k = m]
\]

\[
= \tilde{N}(k, t, i + 1) + [k = m] \frac{m(\mathcal{f}(m, t, i))}{m(1 + \mathcal{f}(\cdot, t))} + \frac{i-1}{m} [k = m]
\]
Preferential Attachment in Graphs with Affinities

\[ N(k, t, i + 1) - c(k, i) \left( \frac{\bar{f}(t, i) + 1}{m(1 + \bar{f}(t))} \right) \left( \frac{mt + i + 1}{1 + \bar{f}(t)} \right) - c(m, t, i) \left( \frac{k = m}{m(1 + \bar{f}(t))} \right) + \frac{i - 1}{m} [k = m] \]

\[ = N(k, t, i + 1) - \left( \frac{\bar{f}(t, i) + 1}{m(1 + \bar{f}(t))} \right) \left( \frac{mt + i + 1}{1 + \bar{f}(t)} \right) - \left( \frac{1}{m} \right) \left( \frac{k = m}{1 + \bar{f}(t)} \right) \]

\[ \times c(k, t, i) \left( \frac{\bar{f}(t, i) + 1}{m(1 + \bar{f}(t))} \right) \left( \frac{mt + i + 1}{1 + \bar{f}(t)} \right) - \left( \frac{1}{m} \right) \left( \frac{k = m}{1 + \bar{f}(t)} \right) \]

\[ = N(k, t, i + 1) + \frac{i - 1}{m} [k = m] - \left( \frac{1}{m} \right) \left( \frac{k = m}{1 + \bar{f}(t)} \right) \]

\[ \times \left( 1 - \left[ k = m \right] \right) \left( k + m(\bar{f}(k, t, i) - 1) \right) + \left( \bar{f}(t, i) + 1 \right) \left( \frac{mt + i + 1}{1 + \bar{f}(t)} \right) - \left( k + m(\bar{f}(k, t, i) - 1) \right) \]

\[ = N(k, t, i + 1) + \frac{i - 1}{m} [k = m] - \left( \frac{1}{m} \right) \left( \frac{k = m}{1 + \bar{f}(t)} \right) \]

\[ \times \left( \frac{c(k, t, i)}{k - 1 + m(\bar{f}(k - 1, t, i) - 1)} \left( \frac{1}{1 + \bar{f}(t)} \right) \cdot (mt + i + 1) - \left( \frac{k = m(\bar{f}(k, t, i) - 1)}{1 + \bar{f}(t)} \right) (mt + i - 1) \right) \]

\[ = \tilde{N}(k - 1, t, i) \left( k - 1 + m(\bar{f}(k - 1, t, i) - 1) \right) + \left( \frac{1}{1 + \bar{f}(t)} \right) \cdot (mt + i - 1) + \tilde{N}(k, t, i) \left( \frac{k = m(\bar{f}(k, t, i) - 1)}{1 + \bar{f}(t)} \right) (mt + i - 1) + \frac{i - 1}{m} [k = m] \]

Note that the last line comes from Eq. (22) and Eq. (23). Now, in a fashion similar to [18] involving manipulations like the ones done above, we have \( \tilde{N}(k, t, i) + [k = m] = O \left( \frac{1}{m} \right) \) finishing the proof of the theorem.

### B Proof of Theorem 3

The proof of the bound on the diameter for PAGA graphs can by obtained by extending the original proof for the standard preferential attachment model from a uniform measure to a non-uniform measure, described by the affinity function \( f \). Here, we ignore the case of having disconnected components in a graph as the graph will be one connected component with high probability.\(^2\) As self loops do not affect the eccentricity of a node and hence the diameter, we simply ignore them in the generation process itself – for the purpose of bounding the diameter.

We begin the proof by noting that for \( m = 1 \) the graph is a tree and for the cases \( m \geq 2 \), which are formed by collapsing the graph \( G^m \) graph, the diameter can only shrink. Next, the diameter of the graph can not be larger than twice the height of this tree, which is equal to the maximal graph distance between vertex 1 and any of the other vertices. So, it is sufficient to bound the height of the tree.

For bounding the tree height, we follow the steps of [48] and outline here the differences.\(^3\) We start with a continuous time branching process, where the rate is given by \( \lambda(t) = d(t) + f(t, t) \). Therefore, the overall transition rate after \( t \) vertices are present (i.e. after \( t - 1 \) births) is given by:

\[ \sum_{j=1}^{t} d(j) + f(t, j) = 2t + t \bar{f}(t, t + 1) \] (24)

Now we can decompose the time \( \tau_t \) as a sum of independent variables, exponentially distributed with parameter \( 2t + t \bar{f}(t, t + 1) \), i.e.

\[ \tau_t = \sum_{j=1}^{t} t_j \]

where \( t_j \sim \text{Exp}(2j + j \bar{f}(t, j + 1)) \). It follows that the mean and variance are bounded by

\[ \mathbb{E}[\tau_t] = \sum_{j=1}^{t} 2j + j \bar{f}(t, j + 1) = O(\log t) \]

\(^2\)In order for a PAGA graph to have a disconnected component, a new super-node \( t + 1 \) has to make \( m \) self loops. This probability asymptotically goes to 0. Note that even in the uniform case this probability is \( \frac{1}{m} \), while the self-loop probability in PAGA is typically smaller than \( \frac{1}{t} \).

\(^3\)To avoid clash of notations, we also redefine the birth times as \( \tau_1, \ldots, \tau_t \) instead of \( t_i \).
\[ \var(\tau_t) = \sum_{j=1}^{t} \frac{1}{2j + jf(\cdot, j + 1)} = O(1) \]

These conditions match the ones required in [48]. Accordingly, the small world property holds for the PAGA model.

**C  Proof of Theorem 4**

Note that at every timestep, there can be at most \( \frac{m(m-1)}{2} \) new triangles added. So the number of triangles \( T(G^t_m) \) is bounded by \( O(n) \). On the other hand, the number of triplets in \( G^t_m \) follows \( \sum_{k=1}^{d_{\max}} N(k, t) \binom{k}{3} \propto \sum_{k=1}^{d_{\max}} t \cdot k \bar{f}(\cdot, t) \).

As the sum \( \sum_{k=1}^{d_{\max}} k = mt \), it is straightforward that \( C(G^t_m) \rightarrow 0 \) unless \( \bar{f}(\cdot, t) > 1 \). We omit the analysis for the cases where \( \bar{f}(\cdot, t) > 1 \) as real graphs follow power law exponent around -2.1 to -2.5 (0.1 \( \leq \bar{f}(\cdot, t) \leq 0.5 \)). □

**D  Spyplot of W**

Figure 6: Spyplot of the matrix \( W \). Blue entries correspond to positive elements; red entries to negative elements.
E  Topics and keywords learned by LDA

We used 50 topics (i.e. $W$ is a $50 \times 50$ matrix) for learning the affinity function $f$. Here, we represent the top 20 keywords for top 10 topics we obtained from LDA.

<table>
<thead>
<tr>
<th>Topic 0</th>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>method</td>
<td>quantization</td>
<td>gauge</td>
<td>theory</td>
<td>gauge</td>
</tr>
<tr>
<td>integral</td>
<td>hamiltonian</td>
<td>invariant</td>
<td>field</td>
<td>magnetic</td>
</tr>
<tr>
<td>path</td>
<td>dirac</td>
<td>invariance</td>
<td>approach</td>
<td>nonabelian</td>
</tr>
<tr>
<td>using</td>
<td>zero</td>
<td>local</td>
<td>based</td>
<td>abelian</td>
</tr>
<tr>
<td>functional</td>
<td>modes</td>
<td>brst</td>
<td>discuss</td>
<td>chernsimons</td>
</tr>
<tr>
<td>regularization</td>
<td>canonical</td>
<td>lagrangian</td>
<td>nonperturbative</td>
<td>electric</td>
</tr>
<tr>
<td>approach</td>
<td>physical</td>
<td>covariant</td>
<td>presented</td>
<td>charge</td>
</tr>
<tr>
<td>formula</td>
<td>lightcone</td>
<td>constraints</td>
<td>properties</td>
<td>theory</td>
</tr>
<tr>
<td>use</td>
<td>coordinates</td>
<td>formulation</td>
<td>new</td>
<td>monopole</td>
</tr>
<tr>
<td>integrals</td>
<td>shown</td>
<td>formalism</td>
<td>methods</td>
<td>su</td>
</tr>
<tr>
<td>expansion</td>
<td>quantized</td>
<td>class</td>
<td>may</td>
<td>monopoles</td>
</tr>
<tr>
<td>used</td>
<td>formalism</td>
<td>first</td>
<td>present</td>
<td>flux</td>
</tr>
<tr>
<td>series</td>
<td>variables</td>
<td>ghost</td>
<td>talk</td>
<td>charged</td>
</tr>
<tr>
<td>procedure</td>
<td>mode</td>
<td>cohomology</td>
<td>version</td>
<td>higgs</td>
</tr>
<tr>
<td>technique</td>
<td>constraint</td>
<td>extended</td>
<td>interpretation</td>
<td>dual</td>
</tr>
<tr>
<td>applied</td>
<td>leads</td>
<td>fields</td>
<td>developed</td>
<td>term</td>
</tr>
<tr>
<td>obtained</td>
<td>operator</td>
<td>lorentz</td>
<td>techniques</td>
<td>vortex</td>
</tr>
<tr>
<td>integration</td>
<td>approach</td>
<td>fixing</td>
<td>provide</td>
<td>selfdual</td>
</tr>
<tr>
<td>expression</td>
<td>constraints</td>
<td>transformations</td>
<td>used</td>
<td>vortices</td>
</tr>
<tr>
<td>formalism</td>
<td>formulation</td>
<td>auxiliary</td>
<td>analysis</td>
<td>yangmills</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topic 5</th>
<th>Topic 6</th>
<th>Topic 7</th>
<th>Topic 8</th>
<th>Topic 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>matrix</td>
<td>coupling</td>
<td>loop</td>
<td>form</td>
</tr>
<tr>
<td>supersymmetric</td>
<td>model</td>
<td>point</td>
<td>theory</td>
<td>general</td>
</tr>
<tr>
<td>supersymmetry</td>
<td>matrices</td>
<td>limit</td>
<td>perturbation</td>
<td>parameters</td>
</tr>
<tr>
<td>supergravity</td>
<td>ansatz</td>
<td>fixed</td>
<td>renormalization</td>
<td>arbitrary</td>
</tr>
<tr>
<td>superconformal</td>
<td>spectral</td>
<td>large</td>
<td>wilson</td>
<td>given</td>
</tr>
<tr>
<td>super</td>
<td>models</td>
<td>strong</td>
<td>finite</td>
<td>explicit</td>
</tr>
<tr>
<td>superspace</td>
<td>relation</td>
<td>flow</td>
<td>perturbative</td>
<td>parameter</td>
</tr>
<tr>
<td>yangmills</td>
<td>limit</td>
<td>points</td>
<td>expansion</td>
<td>q</td>
</tr>
<tr>
<td>multiplet</td>
<td>integrable</td>
<td>constant</td>
<td>one</td>
<td>values</td>
</tr>
<tr>
<td>superfield</td>
<td>spin</td>
<td>group</td>
<td>order</td>
<td>set</td>
</tr>
<tr>
<td>multiplets</td>
<td>smatrix</td>
<td>renormalization</td>
<td>loops</td>
<td>independent</td>
</tr>
<tr>
<td>cal</td>
<td>scattering</td>
<td>weak</td>
<td>non</td>
<td>can</td>
</tr>
<tr>
<td>bps</td>
<td>elements</td>
<td>theory</td>
<td>result</td>
<td>terms</td>
</tr>
<tr>
<td>theories</td>
<td>eigenvalues</td>
<td>infrared</td>
<td>lambda</td>
<td>one</td>
</tr>
<tr>
<td>harmonic</td>
<td>chain</td>
<td>constants</td>
<td>cutoff</td>
<td>structure</td>
</tr>
<tr>
<td>susy</td>
<td>bethe</td>
<td>rg</td>
<td>diagrams</td>
<td>forms</td>
</tr>
<tr>
<td>superfields</td>
<td>corresponding</td>
<td>behavior</td>
<td>divergences</td>
<td>present</td>
</tr>
<tr>
<td>supersymmetries</td>
<td>factors</td>
<td>view</td>
<td>expectation</td>
<td>real</td>
</tr>
<tr>
<td>sym</td>
<td>related</td>
<td>critical</td>
<td>values</td>
<td>allows</td>
</tr>
<tr>
<td>large</td>
<td>representation</td>
<td>expansion</td>
<td>scheme</td>
<td>showns</td>
</tr>
</tbody>
</table>
References


Preferential Attachment in Graphs with Affinities


