Origin Destination Multi Edge Network Package: Additional details on simulations

O. Sagarra[1]

[1]*Departament de Física Fonamental, Universitat de Barcelona, 08028 Barcelona, Spain*

I. GENERATION OF SYNTHETIC NETWORKS ON DIFFERENT ENSEMBLES

This document constitutes the manual for the **Origin Destination Multi Edge network Package** software (ODME), where details on the implementation of the different simulated ensembles are provided. For more details see [1], [2] and [**?**].

The code accompanying the present documents allows for the generation of both directed and undirected multi-edge networks (optionally including or excluding self-loops) in the various ensembles. In the following section a description of the simulation procedure for each ensemble is given. The code for simulating any ensemble can be downloaded from [**?**]. It must be noted that all node related quantities computed are averaged only over the realizations where a given node exists ($s_i \neq 0$), since the probability for a node to have strength 0 is not zero in the GC and Canonical (C) ensembles (albeit rapidly decreases with $\hat{s}$).

II. ENTROPY BASED NULL MODELS

A. Linear constraints

All entropy based models with linear constraints on occupation numbers $\{t_{ij}\}$ are based on the normalized values of p_{ij} which set the expected value of trips between nodes i and j, $\langle t_{ij} \rangle = \hat{T} p_{ij}$ where $\hat{T}$ is the expected number of trips one wants to simulate.

In the case of gravity like models, the average number of events reads,

$$\langle t_{ij} \rangle = x_i y_j e^{-\gamma d_{ij}}$$

which can also be converted to a "power-like" law performing the change $d_{ij} \rightarrow \ln d_{ij}$,

$$\langle t_{ij} \rangle = \frac{x_i y_j}{d_{ij}^{\gamma}}.$$

where $\{x, y\}$ are obtained solving the saddle point equations,

$$\sum_j \langle t_{ij} \rangle = \hat{s}_i^{out} \qquad \sum_i \langle t_{ij} \rangle = \hat{s}_i^{in} \qquad \sum_i j \langle t_{ij} \rangle c_{ij} = \hat{C}.$$

It can also converge to a purely configuration model if $\gamma = 0$, hence,

$$\langle t_{ij} \rangle = x_i y_j = \frac{\hat{s}_i \hat{s}_j}{\hat{T}}.$$

Other options are also possible, like fixing a subset of trips $\mathcal{Q}$ and applying any of the previous model to the complementary set $\bar{\mathcal{Q}}$,

$$\langle t_{ij} \rangle = \begin{cases} \hat{t}_{ij} & ij \in \mathcal{Q} \\ x_i y_j e^{-\gamma c_{ij}} & \in \bar{\mathcal{Q}} \end{cases}.$$

1. Canonical Ensemble

For the canonical ensemble, the statistics of occupation numbers is multinomial with associated probabilities p_{ij} and $\hat{T}$ events. To avoid self-edges one can set $p_{ii} = 0 \,\forall i$. This method has a limited applicability with system size, since the generation of multinomial distributed variables is not independent and requires a large amount of memory, due to the fact that the occupation numbers generated are correlated,

$$\sigma_{t_{ij}, t_{kl}} = \begin{cases} -T p_{ij} p_{kl} & ij \neq kl \\ T p_{ij} (1 - p_{kl}) & ij = kl \end{cases} \tag{1}$$

2. Grand Canonical Ensemble

The grand canonical ensemble can be implemented in two alternative yet equivalent approaches. One can generate a Poisson distributed number τ with mean T and then generate a collection of occupation numbers $\{t_{ij}\}$ using a multinomial distribution of τ trials and probabilities p_{ij} or alternatively one can generate a sequence of L independent Poisson occupation numbers with mean $\langle t_{ij} \rangle = T p_{ij}$. Both approaches are implemented, yet we recommend the latter approach to avoid memory overload problems (in this case the occupation numbers are independent and can be generated accordingly). The complexity of this algorithm scales with the number of possible states L. Note that in this case self-edges can be manually avoided by setting $\langle t_{ii} \rangle = 0$ (or equivalently $p_{ii} = 0$) $\forall i$.

3. Note on averages over existing edges

The results of the randomizer for the average weight of the links as a function of product of strengths or degrees are performed on the existing edges (files *w-s-oi.hist*), hence,

$$\bar{t}(\hat{s}_i, \hat{s}_j) = \langle t | t > 0 \rangle = \frac{\langle t \rangle}{1 - p(0)} = \frac{x_i(\hat{s}_i) y_j(\hat{s}_j)}{1 - e^{-\hat{x}_i \hat{x}_j}}.$$

III. NON-ENTROPY BASED NULL MODELS

We have also implemented the radiation model [3] and the sequential gravity model [4].

A. Radiation Model

We have applied both the stochastic and multinomial versions of the model, introducing also the finite size correction term. All in all, the average number of trips between locations reads,

$$\langle t_{ij} \rangle = \frac{1}{1 - \frac{s_i^{out}}{T}} \frac{s_i^{out} s_j^{in}}{(s_i^{out} + s_{ij})(s_i^{out} + s_j^{in} + s_{ij})} \tag{2}$$

where $s_{ij} = \sum_{j'|d_{ij'} < d_{ij}} s_{j'}^{in}$. The stochastic implementation of the radiation model is much slower than its multinomial counterpart.

B. Sequential gravity model

The sequential gravity model is also implemented in two versions: Bernouilli and multinomial. Note that the Bernouilli model has a slow convergence depending on the form of the strength sequence considered.

---

[1] O. Sagarra, C. J. Pérez Vicente, and A. Díaz-Guilera, Phys. Rev. E **88**, 062806 (2013).
[2] O. Sagarra and F. Font-Clos, arXiv Prepr. arXiv ... , 1 (2014), arXiv:arXiv:1404.3697v1 .
[3] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, Nature **484**, 96 (2012).
[4] M. Lenormand, S. Huet, F. Gargiulo, and G. Deffuant, PLoS One **7**, e45985 (2012), arXiv:1203.5184v1 .