THE STRUCTURE OF PROBABILISTIC NETWORKS

THE MAGICAL MISTERY STOUFFER GROUP

INTRODUCTION

This is a summary of the discussions we had during the lab meeting, with substantially more equations. I don't know about you, but the expressions of *variance* got me over-excited. I haven't put that into the document yet, but I have ran a few tests and it is all matching perfectly.

Each measure defines a *property* on one or several *network units*. These properties can be defined by the unit itself (*direct properties*), or require the association of several units. See Figure 1.

More seriously – we should discuss about a table link each direct property to its unit, and then emerging properties to the groups of units.

METRICS

Throughout this section, we will assume the following notation. $\mathbf{A}$ is a matrix wherein A_{ij} is $\mathrm{P}(ij)$, *i.e.* the probability that species i establishes an interaction with species j. If $\mathbf{A}$ represents a unipartite network (*e.g.* a food web), it is a square matrix and the probabilities of each species interacting with itself. If $\mathbf{A}$ represents a bipartite network (*e.g.* a pollination network), it will most likely not be square. We call S the number of species, and R and C respectively the number of rows and columns. $S = R + C$ in unipartite networks, and $S = R + C$ in bipartite networks. Note that all of the measures defined below can be applied
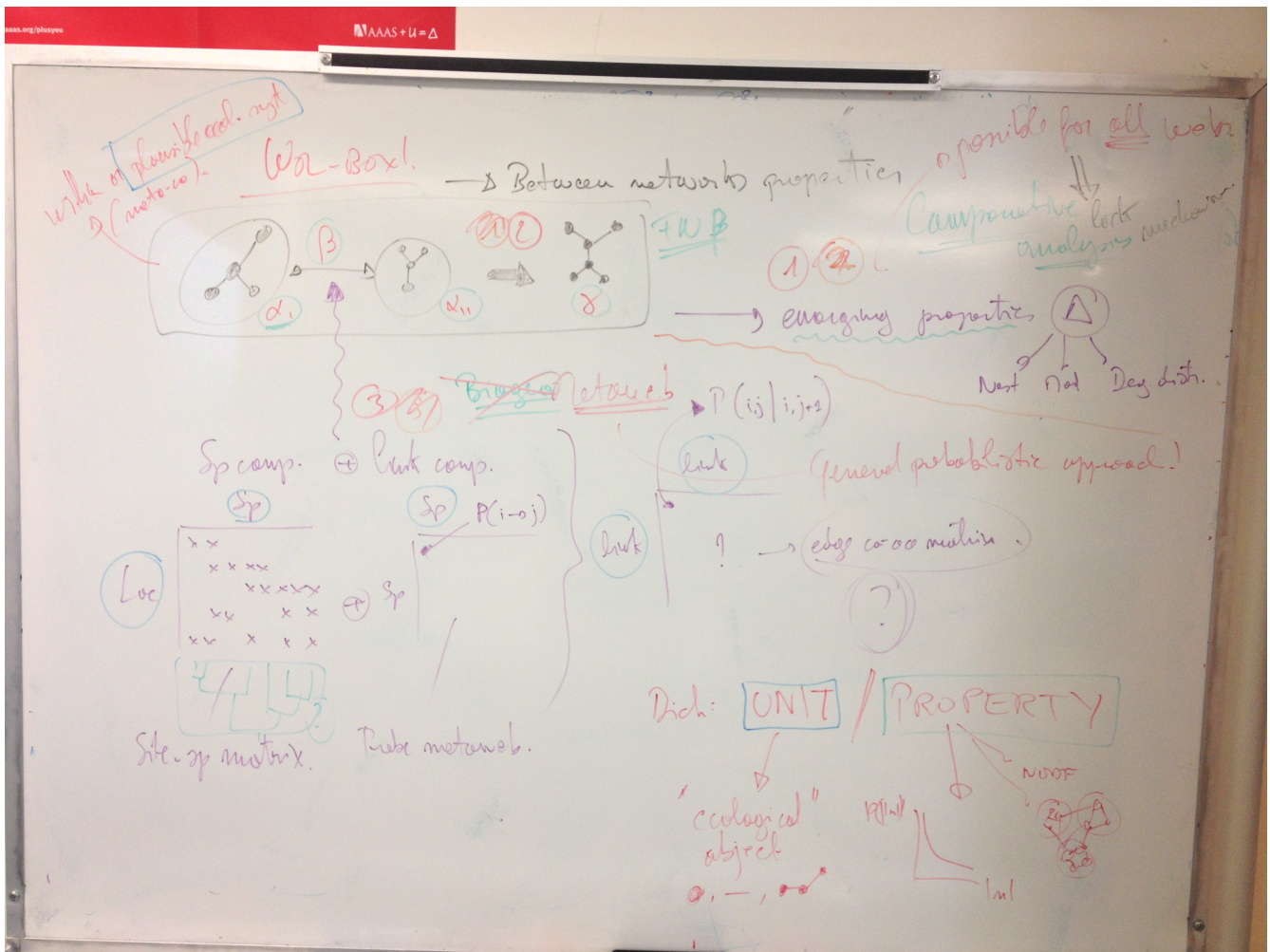
FIGURE 1. Why Tim had his whiteboard privileges revoked.

1. on a bipartite network that has been made unipartite; the unipartite transformation of a bipartite matrix $\mathbf{A}$

2. is the block matrix

(1)
$$\mathbf{B} = \begin{pmatrix} 0_{(R,R)} & \mathbf{A} \\ 0_{(C,R)} & 0_{(C,C)} \end{pmatrix},$$

3. where $0_{(C,R)}$ is a matrix of C rows and R columns filled with 0s, etc.

4. We assume that all interactions are independent (so that $P(ij|kl) = P(ij)P(kl)$ for any species), and can

5. be represented as Bernoulli trials (so that $0 \leq P(ij) \leq 1$). The later condition allows to derive estimates

6. for the *variance* of the measures, since (i) the variance of a single event X_i of probability p is $\mathrm{var}(X) =$

2

1 $p(1-p)$, its expected value is $E(X) = p$, (ii) the variance of additive independent events is the sum of

2 their individual variances, and (iii) the variance of multiplicative independent events is

(2)
$$\text{var}(X_1 X_2 ... X_n) = \prod_i \left(\text{var}(X_i) + [E(X_i)]^2 \right) - \prod_i [E(X_i)]^2$$

3 **Direct properties.**

4 *Connectance and number of interactions.* Connectance is the proportion of realized upon possible inter-

5 actions, defined as $Co = L/(R \times C)$, where L is the total number of interactions. As all interactions in a

6 probabilistic network are assumed to be indpendent, the expected value of L, is

(3)
$$\hat{L} = \sum A_{ij},$$

7 and $\hat{Co} = \hat{L}/(R \times C)$.

8 The variance of the number of interactions is $\text{var}(\hat{L}) = \sum(A_{ij}(1 - A_{ij}))$.

9 *Node degree.* The degree distribution of a network is the distribution of the number of interactions estab-

10 lished and received by each node. The expected degree of species i is

(4)
$$\hat{k}_i = \sum_j (A_{ij} + A_{ji})$$

11 The variance of the degree of each species is $\text{var}(\hat{k}_i) = \sum_j (A_{ij}(1 - A_{ij}) + A_{ji}(1 - A_{ji}))$. Note also that as

12 expected, $\sum \hat{k}_i = 2\hat{L}$.

13 *Average generality and vulnerability.*

14 **Emerging properties.**

15 *Expected trophic level.*

1 *Number of primary producers.* Primary producers, in a food web, are species with no successors, includ-

2 ing themselves. Biologically, they are autotrophic organisms, or organisms whose preys of substrates

3 have been remove from the network. A species is a primary producer if it manages *not* to establish any

4 outgoing interaction, which for species i happens with probability

$$(5) \qquad \prod_j (1 - A_{ij}).$$

5 The number of expected primary producers is therefore the sum of the above across all species:

$$(6) \qquad \hat{PP} = \sum_i \left(\prod_j (1 - A_{ij}) \right).$$

6 The variance in the number of expected primary producers is

$$(7) \qquad \mathrm{var}(\hat{PP}) = \sum_i \left(\prod_j (1 - A_{ij}^2) - \prod_j (1 - A_{ij})^2 \right)$$

7 *Number of top predators.* Top-predators can loosely be defined as species that have no predecessors in

8 the network: they are establishing links with other species, but no species are establishing links with

9 them. Using the same approach than for the number of primary producers, the expected number of

10 top-predators is therefore

$$(8) \qquad \hat{TP} = \sum_i \left(\prod_{j \neq i} (1 - A_{ji}) \right)$$

11 Note that we exclude the self-interactions, as top-predators can, and often do, engage in cannibalism.

12 *Number of species with no interactions.* Predicting the number of species with no interaction (or whether

13 any species will have at least one interaction) is useful to predict whether species will be able to integrate

14 themselves in an existing network, for example.

4

A species has no interactions with probability

$$(9) \qquad \prod_{j \neq i} (1 - A_{ij})(1 - A_{ji})$$

As for the above, the expected number of species with no interactions (*free species*) is the sum of this quantity across all i:

$$(10) \qquad \hat{F}S = \sum_i \prod_{j \neq i} (1 - A_{ij})(1 - A_{ji})$$

The variance of the number of species with no interactions is

$$(11) \qquad \text{var}(\hat{F}S) = \sum_i \left(A_{ij}(1 - A_{ij})A_{ji}(1 - A_{ji}) + A_{ij}(1 - A_{ij})A_{ji}^2 + A_{ji}(1 - A_{ji})A_{ij}^2 \right)$$

Note that from a methodological point of view, this can be a helpful *a priori* measure to determine whether null models of networks will have a lot of species with no interactions, and so will require intensive sampling.

Self-predation. Self-predation (the existence of an interaction of a species onto itself) is only meaningful in unipartite networks. The expected proportion of species with self-loops is very simply defined as $\text{Tr}(\mathbf{A})$, that is, the sum of all diagonal elements. The variance is $\text{Tr}(\mathbf{A} \diamond (1 - \mathbf{A}))$, where $\diamond$ is the element-wise product operation.

Motifs. Motifs are sets of pre-determined interactions between a fixed number of specie [@milo02], such as for example one predator sharing two preys. As there is an arbitrarily large number of motifs, we will illustrate the formulae with only two examples.

The probability that three species form an apparent competition motif (one predator, two preys) where i is the predator, j and k are the preys, is

$$(12) \qquad \mathrm{P}(i, j, k \in \text{app. comp}) = A_{ij}(1 - A_{ji})A_{ik}(1 - A_{ki})(1 - A_{jk})(1 - A_{kj})$$

Similarly, the probability that these three species form an omnivory motif, in which i and j consume k, and i consumes j, is

$$(13) \qquad \mathrm{P}(i, j, k \in \text{omniv.}) = A_{ij}(1 - A_{ji})A_{ik}(1 - A_{ki})A_{jk}(1 - A_{kj})$$

The probability of the number of *any* motif m in a network is given by

$$(14) \qquad \hat{N_\mathrm{m}} = \sum_i \sum_{j \neq i} \sum_{k \neq j} P(i, j, k \in \mathrm{m})$$

It is indeed possible to have an expression of the variance of this value, or of the variance of any three species forming a given motif, but that shit be hardcore yo, and so they are better expressed computationally.

APPLICATIONS

Structure of neutral networks.

Null-model based hypothesis testing.

REFERENCES