

Generating social networks with static and dynamic utility-maximization approaches*

Aldric Labarthe[†] and Yann Kerzreho[‡]

Ecole Normale Supérieure (ENS) Paris-Saclay, Gif-sur-Yvette, 91190 France

(Dated: November 25, 2024)

In this paper, we introduce a conceptual framework that model human social networks as an undirected dot-product graph of independent individuals. Their relationships are only determined by a cost-benefit analysis, *i.e.* by maximizing an objective function at the scale of the individual or of the whole network. On this framework, we build a new artificial network generator in two versions. The first fits within the tradition of artificial network generators by being able to generate similar networks from empirical data. The second relaxes the computational efficiency constraint and implements the same micro-based decision algorithm, but in agent-based simulations with time and fully independent agents. This latter version enables social scientists to perform an in-depth analysis of the consequences of behavioral constraints affecting individuals on the network they form. This point is illustrated by a case study of imperfect information.

I. INTRODUCTION

Existing network models often fall short in accurately representing real social networks, primarily due to limitations in providing a consistent and flexible modeling framework. By consistent we mean the ability to have, at every scale (node or whole network), rules that define agents interactions in a unified way without any *ad hoc* mathematical or computational processes designed to meet pre-established requirements, *e.g.* by (re)-defining edges randomly to check the small world property or to satisfy statistical requirements [1–7]. This excludes the latest advances in graph reproduction using neural networks [8, 9], which are black boxes with no consistent rules on graph formation. By flexible we mean a framework that can cover the whole set of possible human networks and that can be easily modified and extended, *e.g.* towards agent-based modeling.

In the field of Random Geometric Graphs (RGG), nodes are points in a latent space, and their distances from each other allow to decide independently whether they are connected or not (for a review see [10]). The geometry gives good clustering properties for modeling social networks [11] but as a distance must verify the triangular inequality, the small-world property is difficult to achieve [12–15]. The RGG framework is also flexible as different link spaces and functions can be used, and implementing an agent-based model approach is possible [16, 17] with real data [18, 19]. However, the independence assumption between edges is an important limitation because it does not take into account the presence of possible substitution effects between connections. In this paper, we introduce a different model that could remove the independence assumption and avoid the geometric constraints induced by distance. We indeed place ourselves in continuation with dot product graphs models [20]

by using the dot product of the latent space instead of the distance to assess a proximity or compatibility score between individuals.

Following the seminal work of [21] several approaches in the economics literature have tried to model social networks with agent-based postulates, where agents maximize their utility by managing their edges which generate costs and benefits (for an extended review, we refer to [22–24]). This part of the literature uses linear relationship costs and benefits that depend on the graph and not on individuals. [22] stated that the benefits of a relationship are always assumed to be geometrically linked to a closeness measure ([21], or [25] for another definition of closeness). This lack of true link between relationship benefits and specific individuals characteristics leads to contradictions, *e.g.* no individual has a rational incentive in becoming a hub and having many more links than other individuals. Additionally, technical considerations are still quite overlooked in the economics literature [23]. Almost all models heavily rely on agent-based simulations [26] that cannot match the convenience and the efficiency of simulators encountered in the RGG or the machine learning literature.

Our approach can be seen as micro-based, as we define a maximization problem in which each weighted-undirected edge (relationship between individuals) induces a benefit and a cost. Each pair of individuals shares a compatibility which drives up the benefit from the relationship. The weight of an edge represents the intensity of the relationship, accounting for the diversity of social interactions. Our postulates make possible to create graphs by maximizing an objective function. Depending on the scale of the optimization (individual or social), the objective function is the utility function of the individual, or the sum of the utilities of each individual. Therefore, we introduce two artificial network generators: the first uses an analytical optimization process of the social well-being of the population, the latter takes advantage of a more flexible agent-based framework to generate networks that are the result of individual decisions. Hence, our objectives are twofold: using our first model we will introduce a generator that is able to reproduce, with high

* We thank F. Tarissan and F. Bloch for fruitful and useful discussions and J. Randon-Furling for his advice.

[†] aldric.labarthe@ens-paris-saclay.fr (corresponding author)

[‡] yann.kerzreho@ens-paris-saclay.fr

fidelity, every empirical networks (providing they meet some constraints). Secondly, we will use the flexibility of our second generator to study the effects on the network of imperfect information by limiting the agent's scope of action.

We begin by introducing the mathematical framework on which rely both of our generators, and theoretically define our conceptual framework and generators in Section II. The results are then presented in Section III and discussed in Section IV.

II. METHODOLOGY

A. Conceptual Framework

Definition 1 (Individuals space). *We denote as **Individuals space** the product space*

$$\mathcal{I} = (\mathcal{E}, \langle \cdot, \cdot \rangle) \times \mathcal{U}_I$$

with $(\mathcal{E}, \langle \cdot, \cdot \rangle)$ a finite-dimensional euclidean space defined on the vector space \mathcal{E} with $\langle \cdot, \cdot \rangle$ an inner product over \mathbb{R}^+ . Individuals sets are elements of \mathcal{I}^I . No individual should be orthogonal to any other in its first component.

Definition 2 (Individuals graph). *For a given individuals space \mathcal{I} and a given individuals set $\mathcal{J} \in \mathcal{I}^I$, an **individual graph** is a 3-tuple $\mathcal{G}(\mathcal{J}, \alpha, C)$ with α, C two vectors of $(\mathbb{R}_+^*)^{n_I}$ and $(\mathbb{R}_+)^{n_I}$ with $n_I = \binom{I}{2} = \frac{I^2 - I}{2}$:*

- α is a **weight list**.

$$\alpha = (\alpha_{1,2}, \dots, \alpha_{1,I}, \alpha_{2,3}, \dots, \alpha_{2,I}, \dots, \alpha_{I-1,I})^T$$

with $\alpha_{i,j}$ the edge weight between agent i and j

- $\chi(i, j) = \langle P_i, P_j \rangle \mathbb{1}(i \neq j)$ is the **compatibility function**, with $(i, j) \in \mathcal{J}^2$ and P_i the first component of i (in \mathcal{E}).

- C is a **compatibility list**

$$C = (c_{1,2}, \dots, c_{1,I}, c_{2,3}, \dots, c_{2,I}, \dots, c_{I-1,I})^T$$

with $c_{i,j} = \chi(i, j) = \langle P_i, P_j \rangle$.

On each graph is defined $d : \mathcal{I}^2 \rightarrow \mathbb{N}$ the distance between agents i and j : the smallest number of intermediary nonnull edges it takes to reach a vertex from another.

Notation 1. We denote $\forall i \in \llbracket 1..I \rrbracket, \alpha^i = (((\alpha_{i,j})_{i < j \leq I}), ((\alpha_{j,i})_{1 \leq j < i}))$ the $(I-1)$ -dimensional vector of weights associated with all possible edges linking the vertex i with other individuals. α_j^i is the j th-component of α^i . When we write $\alpha_{i,j}$, we consider $\alpha_{\min(i,j), \max(i,j)}$. We use the same notations for C (with C^i and c_j^i).

If we implement a latent space as usual in the RGG field [13], we do not use a standard distance between individuals but rather the dot product of their characteristic

vectors. This choice gives us more freedom, as a standard metric distance would impose compatibilities to satisfy the triangular inequality [12, 13] and would prevent us from finding a suitable compatibility list for every empirical network. In addition, we choose not to provide a probability framework as commonly done in the RGG literature as the focus of this study is essentially placed on empirical networks, and not theoretical ones. However, one can observe that it is perfectly possible to consider a prior distribution for the P variables.

Definition 3 (ALKY utility function). *We consider the set \mathcal{U}_I^{ALKY} of applications $U : [0, 1)^{I-1} \times (\mathbb{R}_+^*)^{I-1} \rightarrow \mathbb{R}$:*

$$\mathcal{U}_I^{ALKY} = \{\kappa > 1, \gamma > 1, \delta > 1, \\ U(\alpha, c) = \sum_{j=1}^{I-1} \left[\kappa \alpha_j c_j - \frac{\alpha_j^\gamma}{1 - \alpha_j^\gamma} \right] - \left(\sum_{j=1}^{I-1} \alpha_j \right)^\delta \}$$

The ALKY function is made up of three parts: a positive one which weights the affinity $(c_{i,j})_{j \leq N}$ by the relationship intensity $(\alpha_{i,j})_{j \leq N}$ and two costs, that convexly penalize intensity, and create substitution effects between relationships. γ is the convexity parameter of relationships' direct cost (and prevents corner solutions). δ enable the substitution effects between relationships. For $\delta > 1$ relations are substitutable: an agent prefers reducing the emphasis on relationships with lower affinity in order to increase the weight on his most successful matches.

Remark 1. All results are stated for a more generic class of utility functions that we define in the supplementary materials.

B. Optimal Graph generation: a social optimization

Definition 4 (Social Optimization Problem). *For a given individuals graph $\mathcal{G}_N(\mathcal{J}, \alpha, C)$, we define the **Social Optimization Problem (SOP)** as:*

$$\begin{cases} \operatorname{argmax}_{\alpha} \mathcal{O}(\alpha, C) = \sum_{i=1}^N U_i(\alpha^i, C^i) \\ \text{s.t.} \quad \begin{cases} \alpha \in [0, 1)^{n_N} \\ \nexists i \in \llbracket 1..N \rrbracket, \forall j \in \llbracket 1..N-1 \rrbracket, \alpha_j^i = 0 \end{cases} \end{cases}$$

With U_i the utility function of agent i , that is its second component in the individuals space.

Theorem 1 (Unicity of the optimal graph). *For any $\mathcal{G}(\mathcal{J}, \alpha, C)$, the Social Optimization Problem has only one unique solution.*

We denote as *Optimal Graph* the unique solutions of the Social Optimization Problem introduced in Definition 4. Using Theorem 2, we know that every network that is associated with a weighted and undirected graph, without any isolated vertex nor loop, can be generated using

our model. In addition, all generated networks verify pairwise stability [21] (see the supplementary materials for a proof).

Theorem 2. *For any individuals set $\mathcal{I} \in (\mathcal{E}, \langle \cdot, \cdot \rangle) \times \mathcal{U}_N$, and weight list α of dimension $n_N = \binom{N}{2}$ satisfying:*

- $\forall a \in \alpha, a \in [0, 1]$
- $\nexists i \in \llbracket 1..N \rrbracket, \forall j \in \llbracket 1..N-1 \rrbracket, \alpha_j^i = 0$
- $\forall i \in \llbracket 1..N \rrbracket, (\mathcal{I})_i = (P_i, U_i), U_i \in \mathcal{U}_I^{ALKY}{}^1$

There exists:

1. *a unique compatibility list \bar{C} such that α is the unique unconstrained solution of the Social Optimization Problem for the individuals graph $\mathcal{G}(\mathcal{I}, \alpha, \bar{C})$*
2. *a set of compatibility lists \underline{C} such that $\forall C \in \underline{C}, \alpha$ is the unique solution of the Social Optimization Problem for the individuals graph $\mathcal{G}(\mathcal{I}, \alpha, C)$.*

$$\underline{C} = \left\{ C \in (\mathbb{R}_+^*)^{n_I}, \quad c_{ij} \in \begin{cases} \{\bar{c}_{ij}\} & \text{if } \alpha_{ij} > 0 \\ (0, \bar{c}_{ij}] & \text{if } \alpha_{ij} = 0 \end{cases} \right\}$$

Moreover, this property will allow us to derive multiple similar synthetic graphs from any real graph, that will share some overall benchmark features, but with different edges and weights. Our methodology is illustrated in Figure 1: we first use Theorem 2 to recover the compatibility matrix associated with a given tuple of parameters (that can be freely chosen) from the real adjacency list. With this tuple of parameters and this compatibility list, the model is always able to regenerate cloned networks (as the solution of the SOP is unique).

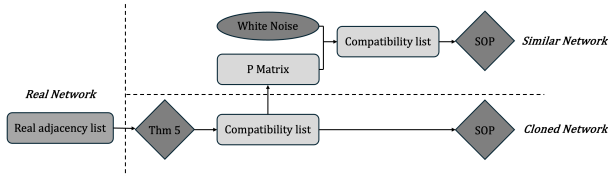


Figure 1. **Methodology to generate similar networks or perfect clones of a real network**

We name P Matrix in figure 1 the matrix that is formed by all individuals' characteristic vectors: the i th row of the P Matrix is the characteristic vector of the individual i (in \mathcal{E}). Several linear algebra methods can be used to recover candidates for these characteristic vectors. A proof of existence of such a P Matrix is provided

in the supplementary materials, and setting the dimension of the latent space to the number of individuals can be taken as a rule of thumb. We can alter this matrix with white-noise processes with different standard deviations. This leads to a new compatibility list and, by applying the SOP to this list, we finally obtain a new graph derived from the real graph.

Theoretically, the resulting network from the Social Optimization problem (SOP) can be seen as a social equilibrium (as usual in the economics literature), that is what would be chosen to maximize the global welfare of the network of individuals. In this approach, all individuals are perfectly substitutable, i. e. it can be optimal to reduce the welfare of one individual to generate a higher welfare for another. One can argue that the model lacks some real properties.

First, the substitution principle does not seem to be appropriate as some individuals can refuse a relationship that does not bring them strictly positive utility gains, and therefore refuse to reduce their welfare for the greater good. Second, both social (Definition 4) and individual (Definition 5) optimization problems are instantaneous and timeless. Third, in both settings, information (in our model, the knowledge of every individual in the network) is assumed to be freely and fully available to all individuals, which is, in the context of human networks, a costly hypothesis, considering that an individual cannot know its affinity with every single individual available. Hence, we will now study networks that are not generated using our Social Optimization Problem (Definition 4) but that are the result of a decentralized optimization problem.

C. Decentralized graph generation: an agent-based model

To tackle these two limitations, we introduce an agent-based model that tries to implement the Individual Optimization Problem (Definition 5) with temporal dynamics and limited information.

Definition 5 (Individual Optimization Problem). *For a given individuals graph $\mathcal{G}_N(\mathcal{I}, \alpha, C)$, we define for all individual $i \in \llbracket 1..N \rrbracket$ the Individual Optimization Problem (IOP) as:*

$$\begin{cases} \operatorname{argmax}_{\alpha^i} U_i(\alpha^i, C^i) \\ \text{s.t. } \alpha^i \in [0, 1]^{I-1} \end{cases}$$

Theorem 3 (Unicity of the individual optimal graph). *For any $\mathcal{G}(\mathcal{I}, \alpha, C)$, and for all individual $i \in \llbracket 1..N \rrbracket$ the Individual Optimization Problem (IOP) has only one unique solution.*

Algorithms 1 and 2 provide details about the implementation: at each time step, each individual is asked (in a random order) to select an action according to the

¹ This condition can be loosen by using the more general definition of utility functions, see the supplementary materials for more details.

gradient of its utility function. This action can be either to create a new relationship (hereinafter, edge), to increase or decrease the weight of an existing edge, or to delete an edge. We introduce the concept of scope, which is a set containing all individuals that a given agent is able to interact with. The scope of an individual i is defined as $\{j : d(j, i) \leq \delta\}$, with δ the scope depth. If the agent wants to increase or create an edge, he has to ask the other individual for his approval, which is again evaluated using the gradient of its own utility evaluated on his own scope. This approach guarantees reciprocity, and implements both constraints of time and imperfect information, as agents are evaluating the gradient of their utility function not on the whole network but only on their scope. Moreover, as an exploration process, agents are selecting their actions using softargmax. A memory of the 3 last actions (and requests) is implemented, preventing individuals to ask again the same others for 3 consecutive periods, and thus incentivizing individuals to explore and select actions that have not the highest gradient value, as it does not take into account possible future gains induced by the scope increase.

Algorithm 1 Individual i select action protocol (simplified)

```

 $\mathcal{S}^i \leftarrow \{j : d(j, i) \leq \delta\}$   $\triangleright$  Get individuals in the scope of  $i$ 
 $C_S^i \leftarrow \{c_{i,j} : j \in \mathcal{S}^i\}$   $\triangleright$  Get the set of compatibilities
 $\alpha_S^i \leftarrow \{\alpha_{i,j} : j \in \mathcal{S}^i\}$   $\triangleright$  Get the set of weights
 $k \leftarrow \text{softargmax}_j |\nabla U^i(\alpha_S^i, C_S^i)|$ 
if  $\nabla U_{(k)}^i > 0$  then
    REQUEST( $k, i, \alpha_{i,k} + \lambda \nabla U_{(k)}^i$ )
else
     $\alpha_{i,k} \leftarrow \alpha_{i,k} + \lambda \nabla U_{(k)}^i$   $\triangleright$  Direct decrease in friendship
    with  $N^k$ 
end if

```

Algorithm 2 Individual k response protocol to i (request function, simplified)

```

function REQUEST( $k, i, a$ )
     $\mathcal{S}^k \leftarrow \{j : d(j, k) \leq \delta\} \cup \{i\}$   $\triangleright$  Get individuals in scope
     $C_S^k \leftarrow \{c_{k,j} : j \in \mathcal{S}^k\}$   $\triangleright$  Get the set of compatibilities
     $\alpha_S^k \leftarrow \{\alpha_{k,j} : j \in \mathcal{S}^k\}$   $\triangleright$  Get the set of weights
    if  $\nabla U_{(i)}^k(\alpha_S^k, C_S^k) > 0$  then
         $\alpha_{i,k} \leftarrow \min(a, \alpha_{i,k} + \lambda \nabla U_{(i)}^k)$ 
        return true
    end if
    return false
end function

```

Theorem 4 provides some conditions on the compatibility structure for the equilibrium to exist, considering that the equilibrium exists if and only if for each pair of individuals (i, j) , they both desire the same $\alpha_{i,j}$ at equilibrium. Under these conditions, the theoretical individual optimal graph are the same as the social opti-

mization problem. This setting will allow us to compute the theoretical optimal graph with our standard generator, and then compare the networks obtained with our decentralized generator (ABM) with respect to it.

Theorem 4 (Invisible hand). *For any $\mathcal{G}(\mathcal{I}, \alpha, C)$ with \mathcal{C} the compatibility matrix defined as $\mathcal{C} = (c_{i,j})_{1 \leq i, j \leq N}$ (\mathcal{C}^i its i th-column), and $(\alpha^i)^*$ the solution vector of the Individual optimization problem (IOP) evaluated on c^i , the compatibility list derived from \mathcal{C}_i , if*

1. $\exists \bar{c} \in \mathbb{R}^N$ s.t. $\forall i \in \llbracket 1..N \rrbracket, \mathcal{C}^i \in \mathcal{S}(\bar{c})$, with $\mathcal{S}(\bar{c})$ the symmetric group of \bar{c}
2. *there exists an ALKY utility function such that $\forall i \in \llbracket 1..N \rrbracket, (\mathcal{S})_i = (P_i, U_i), U_i = U^2$*

Then there exists $\hat{\alpha} \in \mathbb{R}^{n_N}$ such that $\forall i \in \llbracket 1..N \rrbracket, (\alpha^i)^ \in \mathcal{S}(\hat{\alpha})$. Moreover, α^{**} the solution of the associated SOP, verifies:*

$$\forall (i, j) \in \llbracket 1..N \rrbracket^2, \alpha_{i,j}^{**} = (\alpha^{\max\{i,j\}})_{\min\{i,j\}}^*$$

D. Simulation methodology

Hereinafter, we only use homogeneous individuals sets, where all individuals share the same ALKY utility function from Definition 3, with parameters $\gamma = 9, \delta = 2, \kappa = 10$. This should not have any effect on our results as there always exists a compatibility list that produces the same output network with another choice of parameters, as the model has two degree of freedom.

For the Decentralized Graphs, we use Theorem 4 to create compatibility structures in which the Social optimization problem (Definition 4) is equivalent to the decentralized problem, that is the union of Individuals optimization problem (Definition 5). We generate 50 random compatibility structures of 64 individuals that meet the requirements of Theorem 4. Although it is perfectly impossible to cover the whole set of all possible compatibility structures, our sample has been designed to contain heterogeneous networks (the global clustering coefficient is distributed between 0.107 and 0.918, the Gini coefficient between 0.003 and 0.309, the density between 0.127 and 0.921). We use stochastic initialization adjacency lists: $\alpha_{i,j} = (\frac{\omega}{2}, 0; \frac{\iota}{7}, 1 - \frac{\iota}{7})$, with ω the minimum non-negative weight allowed and ι the initialization density. For each compatibility structure, and for each parametrization (ι and δ), we will perform 3 simulations and provide means and standard deviations. All simulations are performed with a fixed duration, but all results are extracted from simulations that can be considered as having reached a limit point: all indicators

² This condition can be loosen by using the more general definition of utility functions, see the supplementary materials for more details.

Table I. **Similar networks generation methodologies assessment on the example of the Train bombing network from Konect** (64 individuals). *All values are means (with a precision of $1e^{-2}$) on a sample of 100 networks generated using the Social Optimization Problem (SOP). Standard deviations across the 100 compatibility structures are in parenthesis with a precision of $1e^{-3}$.*

	Original	WN(σ) on C		WN(σ) on P		
		0.1	0.0001	0.2	0.1	0.05
Vertices	486	325.14	457.76	765.35	489.12	371.59
	(-)	(6.571)	(5.842)	(65.848)	(30.114)	(12.861)
Del. Vertices	-	407.46	319.96	368.62	392.44	399.36
	(-)	(8.394)	(9.268)	(15.019)	(10.28)	(9.304)
New Vertices	-	246.6	291.72	647.98	395.56	284.94
	(-)	(10.657)	(10.483)	(61.387)	(29.745)	(15.605)
Clustering	0.56	0.1	0.24	0.75	0.55	0.29
	(-)	(0.016)	(0.012)	(0.035)	(0.06)	(0.049)
Gini coef.	0.13	0.19	0.23	0.12	0.18	0.21
	(-)	(0.013)	(0.008)	(0.013)	(0.012)	(0.012)
Density	0.12	0.08	0.11	0.19	0.12	0.09
	(-)	(0.002)	(0.001)	(0.016)	(0.007)	(0.003)
Avg. distance	0.63	1.25	0.5	1.99	1.66	1.39
	(-)	(0.065)	(0.02)	(0.348)	(0.182)	(0.122)
Assortativity	0.03	-0.05	-0.11	0.15	0.09	-0.18
	(-)	(0.057)	(0.035)	(0.13)	(0.12)	(0.086)
Avg. degree	7.59	5.08	7.15	11.96	7.64	5.81
	(-)	(0.103)	(0.091)	(1.029)	(0.471)	(0.201)
Var. degree	37.99	7.47	30.19	43.19	16.2	10.67
	(-)	(0.508)	(0.876)	(6.537)	(2.699)	(0.819)
S.d. Comp.	0.28	0.3	0.28	2.59	0.74	0.35
	(-)	(0.002)	(0)	(0.433)	(0.111)	(0.02)
Cor. ¹ Comp.	0.9	0.73	0.9	0.05	0.21	0.55
	(-)	(0.005)	(0)	(0.02)	(0.042)	(0.055)
S.d. Cor. ²	0.14	0.12	0.14	0.89	0.69	0.36
	(-)	(0.003)	(0)	(0.027)	(0.043)	(0.051)

¹ Cor. denotes the Pearson coefficient between individuals' compatibility lists (for agent i , $c^i = (c_{1,i}, \dots, c_{N,i})$)

² Sd. Cor. is the standard dispersion of the Pearson coefficients.

(global clustering coefficient, density, Gini, ...) do not exhibit any statistically significant variations in the last 5% time, or the round-by-round simulation has registered no accepted actions in the last 50 rounds.

III. RESULTS

A. Similar networks generation

We here demonstrate on an example the ability of the model not only to regenerate the network, but to enable the production of very similar networks in their overall benchmark properties, but with different edges and weights. We use as an example the real network named "Moreno Train bombing" [27] available in the Konect database [28]. This network is made of "contacts between suspected terrorists involved in the train bombing of Madrid on March 11, 2004 as reconstructed from newspapers."

One naive approach would be to try to directly alter the compatibility list. Nonetheless the model's response to compatibility list is not continuous. Table I

features the results of the implementation of a normally distributed white-noise of standard-deviation of 0.1 and 0.0001 over two samples of 100 networks. Even with such small changes in the compatibility list, the number of edges drops significantly which dramatically alters the clustering coefficient making networks that cannot be considered as "similar" with the original one. One possible explanation can lie in the white-noise deteriorating the correlation between columns of the compatibility matrix. Moreover, this high sensitivity to small changes in the compatibility list should be seen with respect to the high robustness of the main benchmark characteristics of the original graph: when removing 5% of the vertices the clustering coefficient follows a standard distribution $\mathcal{N}(0.5055, 0.0148)$ (sample of 1000 trials, Shapiro test p-value of 0.04) whereas the original clustering coefficient was 0.5610. Similar results can be achieved with different methods: when removing 10% of individuals and considering the resulting sub-graphs, we obtain an average global clustering coefficient of 0.5510 (standard deviation of 0.026 over 1000 observations), illustrating a high robustness of the network. Due to the high sensibility of the model to the compatibility list, this method does not appear relevant.

Our suggested methodology tries to take into account this criticism: the compatibility matrix has columns that are correlated because the compatibility list is formed by scalar-products of the individuals' characteristic vector. Hence, we suggest not directly altering the results of these scalar-products, but the the individuals' characteristic vector.

Results are shown in Table I. Rather than trying to infer a specific relationship between dispersion of the white-noise process and main benchmark characteristics (a relationship that would risk to be specific to this example), we only observe that the behavior of the output is much more continuous with the dispersion of the white-noise process than what we observed when we altered directly the compatibility list. Moreover, we find that, for this example, the dispersion $\sigma = 0.1$ gives the best results, with networks that are very similar: the global clustering coefficient, the density, the average degree and the number of relations are very close to what characterizes the original graph. The networks obtained are also sufficiently different from the original to not to be considered as cloned: around 80% of the edges that exist in the new networks did not exist in the original one.

B. Decentralized optimization and agent-based model

We will now try to measure the distance between the decentralized equilibrium and the theoretical one (SOP). We first challenge the robustness of the generator to its input parameters, and then try to assess the effects of the frictions on the equilibrium.

First, despite the randomness of the generation

Table II. **Random Effects regression estimates of the heterogeneity of the effect of the scope dimension (δ) on the deviation of the equilibrium of the ABM from the theoretical solution (SOP).** In models (1) to (6), \varkappa is the clustering coefficient of the optimal graph (solution of the SOP), in model (7) it is the Gini coefficient of the same graph (*sample of 50 compatibility structures, simulated with $\iota = 4$*).

	Dependent variable:						
	Clustering (1)	Distance (2)	Avg. degree (3)	Avg. Utility (4)	S.d. Utility (5)	Density (6)	Gini (7)
$\delta = 0$	-0.424*** (0.067)	-0.417*** (0.076)	-0.482*** (0.023)	-0.672*** (0.014)	435,973 (392,293)	-0.482*** (0.023)	-9.723*** (1.571)
$\delta = 1$	0.582*** (0.054)	0.441*** (0.062)	-0.279*** (0.019)	-0.072*** (0.011)	962,152*** (318,603)	-0.279*** (0.019)	4.922*** (1.263)
$\delta = 2$	-0.207*** (0.054)	0.084 (0.062)	-0.301*** (0.019)	-0.103*** (0.011)	791,552** (318,603)	-0.301*** (0.019)	1.479 (1.263)
$\delta = 3$	-0.301*** (0.054)	0.156** (0.062)	-0.346*** (0.019)	-0.130*** (0.011)	741,601** (318,603)	-0.346*** (0.019)	0.347 (1.263)
$\delta = 4$	-0.322*** (0.067)	0.249*** (0.076)	-0.364*** (0.023)	-0.141*** (0.014)	899,695** (392,293)	-0.364*** (0.023)	-3.700** (1.571)
\varkappa	-0.011 (0.113)	0.468* (0.278)	-0.152** (0.077)	0.154*** (0.017)	-647,334 (1,692,293)	-0.152** (0.077)	-53.950*** (9.179)
$\varkappa \mathbb{I}_{[\delta=0]}$	-0.318** (0.140)	0.765*** (0.160)	-0.254*** (0.049)	-0.325*** (0.029)	-335,625 (818,832)	-0.254*** (0.049)	53.950*** (10.511)
$\varkappa \mathbb{I}_{[\delta=1]}$	-1.463*** (0.114)	-0.649*** (0.130)	-0.008 (0.040)	-0.173*** (0.023)	-763,892 (668,573)	-0.008 (0.040)	-26.215*** (8.582)
$\varkappa \mathbb{I}_{[\delta=2]}$	-0.196* (0.114)	0.144 (0.130)	-0.060 (0.040)	-0.192*** (0.023)	-640,581 (668,573)	-0.060 (0.040)	-8.289 (8.582)
$\varkappa \mathbb{I}_{[\delta=3]}$	-0.051 (0.114)	0.088 (0.130)	0.006 (0.040)	-0.159*** (0.023)	-499,933 (668,573)	0.006 (0.040)	-2.254 (8.582)
$\varkappa \mathbb{I}_{[\delta=4]}$	-0.023 (0.140)	-0.043 (0.160)	0.046 (0.049)	-0.146*** (0.029)	-644,439 (818,832)	0.046 (0.049)	22.598** (10.511)
Cst.	0.730*** (0.054)	1.352*** (0.133)	0.838*** (0.037)	0.883*** (0.008)	591,077 (807,459)	0.838*** (0.037)	9.723*** (1.378)
Adj. R^2	0.685	0.550	0.930	0.983	0.065	0.930	0.244
F Stat.	1,098***	623***	6,680***	29,537***	47***	6,680***	173***

Note:

500 observations, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

seed, the intra-dispersion (the dispersion in results between the three runs associated with the same compatibility list) is negligible with respect to the dispersion observed between compatibility structures (always more than 10 times lower), with most indicators having dispersion of less than 2% (max: 9% in our sample) of their mean value. We observe a quadratic relation between the initialization density and the deviation from the theoretical graph for the clustering coefficient and the density, and long lasting effects on the output that need to be taken into account for later comparisons. Transitioning to a decentralized generation process reduces the maximum value reached by the objective function at equilibrium (optimality cost): the reduction of the average utility received by individuals is ranging from 6% to 1% for $\iota \in \{4, 16, 32, 48\}$ (the seed density). This illustrates that the round-by-round decentralized process is able to find an equilibrium that is very close to the theoretical optimum.

Second, implementing scope limitations always re-

duces the utility reached by agents (objective function), while decreasing the clustering coefficient, the average degree and the density (Table II). The effect is significantly increasing with δ , the scope depth. The utility dispersion is higher with scope limitations, but the impact of δ is unclear. Among all random effects models with clustered errors for the three runs, the average utility appears to be the most prominent and significant relationship, with very high significance of all regressors and a high adjusted R^2 (0.979), taking into account the effect of the initialization seed.

When controlling by the clustering coefficient (or the Gini coefficient) that has been observed in the SOP solution, clustering, distance, average utility and Gini coefficients are shown to be altered differently by the scope limitations, depending on the type of theoretical network. First, these estimates make the case $\delta = 0$ much more understandable: among all scope limitations, it is the most sensitive to the theoretical clustering of the graph, illustrating the inability of the agent-based model to gener-

ate highly clustered graphs with such scope limitations. Second, except for the Gini coefficient, the higher is δ (that is the lower are the scope limitations), the less are its effects dependent on the theoretical clustering coefficient, suggesting that this dependency would essentially matter for high scope limitations ($\delta \leq 1$). Last but not least, the optimality cost appears to be highly dependent of the theoretical optimal network clustering coefficient, with lower clustered networks being much more resilient to frictions. The cross-effect of scope limitations and theoretical clustering is also steadily decreasing with δ .

Overall, scope limitations have an ambiguous effect. As widely experimented in the Economics literature, imperfect information is reducing the quality of the optimum (the objective function reaches a lower value, here the total utility is lower). Nonetheless, this reduction of the quality of the equilibrium does not appear to be linearly decreasing with the amount of information accessible to agents, suggesting a quadratic relation with information. The network benchmark properties are affected heterogeneously, but all resulting networks share a common reduction of the total number of edges and nodes' degree, inducing a reduction of the global density, an increase in the mean distance and an alteration of the global clustering coefficient. All compatibility structures exhibits similar reactions to frictions, but less clustered networks have proved slightly more resilient, especially when frictions are very high.

IV. DISCUSSION

In the sub-field of graph reproduction, our results with the Optimal Graph generator (from the SOP, Definition 4) have illustrated that not only it is possible to use our objective-maximization framework to model human networks, but also to generate similar networks from an original one [29–31]. This suggests that it is perfectly possible to conciliate reproducibility and fidelity with the original graph, with interpretability of the generation process.

Moreover, our micro-based framework could be extended to numerically estimate the utility function parameters or compatibilities between individuals using real graphs. This would help link pure-graph features (*e.g.*, clustering coefficient, degree distribution) to the behavior of network agents. However, this requires addressing two degrees of freedom: the compatibility matrix and the utility function parameters. Two approaches could be explored: inferring parameters from a compatibility distribution or inferring compatibilities from fixed parameters. The first approach standardizes compatibilities by settings the parameters of the utility function, allowing comparisons. However, as determining these distributions and effectively setting the utility functions seems complex, the second approach may be preferred. It involves fixing parameters to derive compatibilities, enabling comparisons if graphs are supposed to share the

same utility parameters.

The question of utility function selection is important for several reasons. First even if we can generalize our results for a class of functions (cf. supplementary materials), one can argue that we have no guarantee that the agent-based model would present similar properties with a different function. If our function is more complex than what was previously experimented [22, 23] making relationships both substitutable and not linear in their costs, several other determinants could be attached. The centrality implemented in [21] could be added by making the advantage of having a relationship with two individuals depend on the fact that these two individuals are themselves related. In addition, the choice of the utility function raises the question of the computational efficiency of both generators. The downside of our fully micro-based approach is that it is comparatively less efficient than other generators that do not care for consistency of agents behaviors at each scale. Some later developments could therefore explore new utility functions that could induce computational efficiency gains, allowing the generator to handle higher dimensional networks.

Furthermore, this work strengthens the need for further developments on the consequences of the latent space geometry. Several works including [14, 15] have tried to show that the degree distribution can be a power law distribution only using non-euclidean geometry, but the theoretical implications of using a type of geometry rather than another remain an open question today. Our choice of using a scalar product as a compatibility measure can be seen as implementing a non-euclidean latent space with $d^{\text{latent}}(x, y) = (\chi(x, y))^{-1}$ (for $x \neq y$, 0 elsewhere). This conic non-linear geometry has allowed us to recover similar results as the one found with hyperbolic geometry, without clear explanation.

Finally, our approach paves the way for studying the consequences of real world imperfections on the shape of networks with our generators. Indeed, the Optimal Graph generator can be used as a proxy for what would have been obtained with independent and rational agents in conditions given by Theorem 4. On the other hand, the agent-based model can be used to obtain estimates of the impact of real-life constraints such as geographic or normative ones.

CONCLUSION

Overall, our main contribution has been to introduce a mathematical framework able to conciliate three distinct objectives: fidelity to empirical network when generating from real-world data, consistency at every scale in the generation process with fully individual-based postulates, and versatility in the implementation allowing for the use of agent-based round-by-round simulations. Indeed, our Optimal graph generator has been proven to be able to regenerate any real-life network (that satisfy very loose criteria) without leaving aside our

agent-based assumptions. Moreover, we have illustrated on an example that our framework was even able to generate very similar artificial networks from an empirical one.

Our micro-based consistency has enabled us to build an agent-based simulator with round-by-round rules, with approximately zero optimality cost, justifying the relevance of the coexistence of the two models: if the agent-based model allows to test hypothesis and real-world constrains, the analytical Optimal graph model offers a quicker and a more versatile generator for users that would not care for such hypothesis tests. This has enabled us to challenge the effect of two real-world con-

strains, time and limited information, on the resulted network: for the limited information constrain, highly clustered networks and populations with highly dispersed affinity scores have proven to be less resilient.

CODE AND DATA AVAILABILITY

Our implementation of the generators is available in C++ on the GitHub repository where both the source code and the executable can be found under the GNU General Public License v3.0. All simulations data are also provided, with the R script that enabled us to create all tables and to compute all statistical estimations.

-
- [1] D. J. Watts and S. H. Strogatz, Collective dynamics of ‘small-world’ networks, *nature* **393**, 440 (1998).
 - [2] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *science* **286**, 509 (1999).
 - [3] M. O. Jackson and B. W. Rogers, Meeting strangers and friends of friends: How random are social networks?, *American Economic Review* **97**, 890 (2007).
 - [4] L. Hamill and N. Gilbert, Social circles: A simple structure for agent-based social network models, *Journal of Artificial Societies and Social Simulation* **12**, 3 (2009).
 - [5] M.-D. Pham, P. Boncz, and O. Erling, S3g2: A scalable structure-correlated social graph generator, in *Technology Conference on Performance Evaluation and Benchmarking* (Springer, 2012) pp. 156–172.
 - [6] O. Erling, A. Averbuch, J. Larriba-Pey, H. Chafi, A. Gubichev, A. Prat, M.-D. Pham, and P. Boncz, The ldbc social network benchmark: Interactive workload, in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, SIGMOD ’15 (Association for Computing Machinery, New York, NY, USA, 2015) p. 619–630.
 - [7] C. Zhuge, C. Shao, and B. Wei, An agent-based spatial urban social network generator: A case study of beijing, china, *Journal of computational science* **29**, 46 (2018).
 - [8] R. Liao, Y. Li, Y. Song, S. Wang, W. Hamilton, D. K. Duvenaud, R. Urtasun, and R. Zemel, Efficient graph generation with graph recurrent attention networks, *Advances in neural information processing systems* **32** (2019).
 - [9] Y. Zhu, Y. Du, Y. Wang, Y. Xu, J. Zhang, Q. Liu, and S. Wu, A survey on deep graph generation: Methods and applications, in *Learning on Graphs Conference* (PMLR, 2022) pp. 47–1.
 - [10] Q. Duchemin and Y. De Castro, Random geometric graph: Some recent developments and perspectives, *High Dimensional Probability IX: The Ethereal Volume*, 347 (2023).
 - [11] D. Krioukov, Clustering implies geometry in networks, *Physical review letters* **116**, 208302 (2016).
 - [12] S. Talaga and A. Nowak, Homophily as a process generating social networks: insights from social distance attachment model, *arXiv preprint arXiv:1907.07055* (2019).
 - [13] R. Iijima, R. Iijima, Y. Kamada, and Y. Kamada, Social distance and network structures: Social distance and network structures, *Theoretical Economics* 10.3982/te1873 (2017).
 - [14] M. Papamichalis, K. Turnbull, S. Lunagomez, and E. Airolidi, Latent space network modelling with hyperbolic and spherical geometries, 2109.03343.
 - [15] S. K. Saha and T. Gosswami, A study of triangle inequality violations in social network clustering, *Journal of Computer and Communications* **12**, 67 (2024), num Pages: 10 Number: 01.
 - [16] M. C. González, P. G. Lind, and H. J. Herrmann, Networks based on collisions among mobile agents, *Physica D: Nonlinear Phenomena* **224**, 137 (2006).
 - [17] R. De Caux, C. Smith, D. Kniveton, R. Black, and A. Philippides, Dynamic, small-world social network generation through local agent interactions, *Complexity* **19**, 44 (2014).
 - [18] O. Klopp, A. B. Tsybakov, and N. Verzelen, Oracle inequalities for network models and sparse graphon estimation, (2017).
 - [19] O. Klopp and N. Verzelen, Optimal graphon estimation in cut distance, *Probability Theory and Related Fields* **174**, 1033 (2019).
 - [20] S. J. Young and E. Scheinerman, Directed random dot product graphs, *Internet Mathematics* **5**, 91, publisher: A K Peters, Ltd.
 - [21] M. O. Jackson, M. O. Jackson, A. Wolinsky, and A. Wolinsky, A strategic model of social and economic networks, *Journal of Economic Theory* 10.1006/jeth.1996.0108 (1996).
 - [22] T. Hellmann and M. Staudigl, Evolution of social networks, *European Journal of Operational Research* **234**, 583 (2014).
 - [23] M. O. Jackson, The study of social networks in economics, The missing links: Formation and decay of economic networks **76**, 210 (2007).
 - [24] M. O. Jackson, An overview of social networks and economic applications, *Handbook of social economics* **1**, 511 (2011).
 - [25] P. Bonacich, Power and centrality: A family of measures, *American journal of sociology* **92**, 1170 (1987).
 - [26] L. Tesfatsion, A trade network game with endogenous partner selection, in *Computational approaches to economic problems* (Springer, 1997) pp. 249–269.
 - [27] Train bombing network dataset – KONECT (2017).

- [28] J. Kunegis, KONECT – The Koblenz Network Collection, in *Proc. Int. Conf. on World Wide Web Companion* (2013) pp. 1343–1350.
- [29] A. M. Ali, H. Alvari, A. Hajibagheri, K. Lakkaraju, and G. Sukthankar, Synthetic generators for cloning social network data, *Proceedings of SocInfo* (2014).
- [30] D. F. Nettleton, A synthetic data generator for online social network graphs, *Social Network Analysis and Mining* **6**, 1 (2016).
- [31] S. Johnson, Epidemic modelling requires knowledge of the social network, *Journal of Physics: Complexity* **5**, 01LT01 (2024).

Appendix A: Theoretical and mathematical Framework

1. Fundamental definitions

Definition 6 (Utility function). For a given $I \in \mathbb{N}^*$, a utility function $U(\alpha, C)$ is an element of \mathcal{U}_I , the set of applications $U : \mathcal{R}^{I-1} \times (\mathbb{R}_+^*)^{I-1} \rightarrow \mathbb{R}$ with $[0, 1] \subseteq \mathcal{R} \subseteq \mathbb{R}$, and \mathcal{R}^{I-1} convex. Every element $U : (\alpha, C) \mapsto \mathbb{R}$ of \mathcal{U}_I verifies:

1. $\forall C \in (\mathbb{R}_+^*)^{I-1}, U(0, C) = 0$, and $\frac{\partial U}{\partial \alpha_i}(0) > 0$
2. with H bijective and increasing in its second component,

$$\exists \begin{cases} H : \mathbb{R}_+^* \times \mathcal{R} \rightarrow \mathbb{R}, & H \in \mathcal{C}^1 \\ S : \mathcal{R}^{I-1} \rightarrow \mathbb{R}, & S \in \mathcal{C}^1 \end{cases} \quad s.t. \quad \forall i, \frac{\partial U}{\partial \alpha_i} = H(c_i, \alpha_i) + S(\alpha)$$

3.

$$\exists \begin{cases} h : \mathbb{R}_+^* \times \mathcal{R} \rightarrow \mathbb{R}, & h \in \mathcal{C}^0 \\ s : \mathcal{R}^{I-1} \rightarrow \mathbb{R}, & s \in \mathcal{C}^0 \end{cases} \quad s.t. \quad \forall j \neq k, \quad \frac{\partial^2 U}{\partial \alpha_j \partial \alpha_k} = s(\alpha), \quad \frac{\partial^2 U}{\partial \alpha_j^2} = h(c_j, \alpha_j) + s(\alpha)$$

$$\text{with} \begin{cases} S(\alpha) = 0 \vee s(\alpha) = 0 \Leftrightarrow \alpha = 0 \\ S(\alpha) = 0 \vee s(\alpha) = 0 \Rightarrow \forall j \leq I-1, h(c_j, \alpha_j) = 0 \wedge H(c_j, \alpha_j) = 0 \\ \forall \mathcal{A} \in \mathcal{R}^{I-1}, \exists \bar{s} \in \mathbb{R}, \forall a \in \mathcal{S}(\mathcal{A}), S(a) = \bar{s} \\ \forall \alpha, \forall i, h(\alpha_i, c_i) \leq 0, s(\alpha) \leq 0, S(\alpha) \leq 0 \end{cases}$$

Lemma 1. All Utility functions are strictly concave in their first argument on \mathcal{R}^{I-1} .

Proof (Lemma 1). Let's denote the Hessian of $U \in \mathcal{U}$ and take $(C, \alpha) \in (\mathbb{R}_+^*)^{I-1} \times \mathcal{R}^{I-1}$ with $\alpha \neq 0$,

$$\begin{aligned} \mathbf{H} &= \begin{pmatrix} h(c_1, \alpha_1) + s(\alpha) & s(\alpha) & \dots & s(\alpha) \\ s(\alpha) & h(c_2, \alpha_2) + s(\alpha) & \ddots & \vdots \\ \vdots & \ddots & \ddots & s(\alpha) \\ s(\alpha) & \dots & s(\alpha) & h(c_{I-1}, \alpha_{I-1}) + s(\alpha) \end{pmatrix} \\ \Rightarrow \det(\mathbf{H}) &= \begin{vmatrix} h(c_1, \alpha_1) + s(\alpha) & s(\alpha) & \dots & s(\alpha) \\ s(\alpha) & h(c_2, \alpha_2) + s(\alpha) & \ddots & \vdots \\ \vdots & \ddots & \ddots & s(\alpha) \\ s(\alpha) & \dots & s(\alpha) & h(c_{I-1}, \alpha_{I-1}) + s(\alpha) \end{vmatrix} \\ \Rightarrow \det(\mathbf{H}) &= \begin{vmatrix} h(c_1, \alpha_1) & 0 & \dots & 0 \\ 0 & h(c_2, \alpha_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & h(c_{I-1}, \alpha_{I-1}) \end{vmatrix} \end{aligned}$$

As h and s have been assumed to be strictly negative for $\alpha \neq 0$, we have that \mathbf{H} is negative definite on $\mathcal{R}^{I-1} \setminus \{0\}$. Let's prove strict concavity over \mathcal{R}^{I-1} . We should have $\forall (x, y) \in (\mathcal{R}^{I-1})^2, x \neq y, \forall \lambda \in (0, 1), U(\lambda x + (1 - \lambda)y, C) < \lambda U(x, C) + (1 - \lambda)U(y, C)$. We can distinguish two cases:

- $x \neq 0, y \neq 0$, the standard second characterization of strict concavity provides the proof as \mathbf{H} is negative definite on the domain of x and y .
- $x = 0$ or $y = 0$ (let's take $x = 0$): $U(\lambda x + (1 - \lambda)y, C) = U((1 - \lambda)y, C)$. As we have strict concavity everywhere but in 0 and as $U(0, C) = 0$ we have

$$U((1 - \lambda)y, C) < (1 - \lambda)U(y, C) \Rightarrow U(\lambda x + (1 - \lambda)y, C) < (1 - \lambda)U(y, C) = \lambda U(x, C) + (1 - \lambda)U(y, C)$$

□

Definition 7 (Individuals space). We denote as **Individuals space** the product space

$$\mathcal{I} = (\mathcal{E}, \langle \cdot, \cdot \rangle) \times \mathcal{U}_I$$

with $(\mathcal{E}, \langle \cdot, \cdot \rangle)$ a finite-dimensional euclidean space defined on the vector space \mathcal{E} with $\langle \cdot, \cdot \rangle$ an inner product over \mathbb{R}^+ . Individuals sets are elements of \mathcal{I}^I . No individual should be orthogonal to any other in its first component.

Definition 8 (Individuals graph). For a given individuals space \mathcal{I} and a given individuals set $\mathcal{J} \in \mathcal{I}^I$, an **individual graph** is a 3-tuple $\mathcal{G}(\mathcal{J}, \alpha, C)$ with α, C two vectors of $(\mathbb{R}_+^*)^{n_I}$ and $(\mathbb{R}_+)^{n_I}$ with $n_I = \binom{I}{2} = \frac{I^2 - I}{2}$:

- α is a **weight list**.

$$\alpha = (\alpha_{1,2}, \dots, \alpha_{1,I}, \alpha_{2,3}, \dots, \alpha_{2,I}, \dots, \alpha_{I-1,I})^T$$

with $\alpha_{i,j}$ the edge weight between agent i and j

- $\chi(i, j) = \langle P_i, P_j \rangle \mathbb{1}(i \neq j)$ is the **compatibility function**, with $(i, j) \in \mathcal{J}^2$ and P_i the first component of i (in \mathcal{E}).
- C is a **compatibility list**

$$C = (c_{1,2}, \dots, c_{1,I}, c_{2,3}, \dots, c_{2,I}, \dots, c_{I-1,I})^T$$

with $c_{i,j} = \chi(i, j) = \langle P_i, P_j \rangle$.

On each graph is defined $d : \mathcal{I}^2 \rightarrow \mathbb{N}$ a distance between agents i and j : the smallest number of intermediary non-null edges it takes to reach a vertex from another.

Remark 2. In a Individuals graph, the individual utility of agent i , $U_i \in \mathcal{U}_I$ is evaluated on α^i and on C^i , that is on the vectors of weights and compatibilities with respect to agent i .

Notation 2. We denote $\forall i \in \llbracket 1..I \rrbracket, \alpha^i = (((\alpha_{i,j})_{i \leq j \leq I}), ((\alpha_{j,i})_{1 \leq j < i}))$ the vector of all vertices liked with the vertex i of dimension $I - 1$. α_j^i is the j th-component of α^i . When we write $\alpha_{i,j}$, we consider $\alpha_{\min(i,j), \max(i,j)}$. We use the same notations for C (with C^i and c_j^i).

2. A network optimization problem

Definition 9 (Individual Optimization Problem). For a given individuals graph $\mathcal{G}_N(\mathcal{J}, \alpha, C)$, we define for all individual $i \in \llbracket 1..N \rrbracket$ the Individual Optimization Problem (IOP) as:

$$\begin{cases} \operatorname{argmax}_{\alpha^i} U_i(\alpha^i, C^i) \\ \text{s.t. } \alpha^i \in [0, 1]^{I-1} \end{cases} \quad (\text{A1})$$

Theorem 5 (Unicity of the individual optimal graph). For any $\mathcal{G}(\mathcal{J}, \alpha, C)$, and for all individual $i \in \llbracket 1..N \rrbracket$ the Individual Optimization Problem (IOP) has only one unique solution.

Proof (Theorem 5). By Lemma 1, we have that U_i is strictly concave in its first argument on \mathcal{R}^{I-1} , and on $[0, 1]^{I-1}$ as it is a convex subset of \mathcal{R}^{I-1} . As constraints are qualified and convex, we have that the IOP has one unique solution. \square

Definition 10 (Social Optimization Problem). For a given individuals graph $\mathcal{G}_N(\mathcal{J}, \alpha, C)$, we define the Social Optimization Problem (SOP) as:

$$\begin{cases} \operatorname{argmax}_{\alpha} \mathcal{O}(\alpha, C) = \sum_{i=1}^N U_i(\alpha^i, C^i) \\ \text{s.t. } \begin{cases} \alpha \in [0, 1]^{n_N} \\ \nexists i \in \llbracket 1..N \rrbracket, \forall j \in \llbracket 1..N-1 \rrbracket, \alpha_j^i = 0 \end{cases} \end{cases} \quad (\text{A2})$$

With U_i the utility function of agent i , that is its second component in the individuals space.

Lemma 2 (Strict concavity of the objective). For a given individuals graph $\mathcal{G}_I(\mathcal{J}, \alpha, C)$, $\mathcal{O}(C, \alpha) = \sum_{i=1}^I U_i(C^i, \alpha^i)$ is strictly concave on the feasible set $\{\alpha \in \mathcal{R}^{I-1}, [\alpha \in [0, 1]^{n_I}] \wedge [\nexists i \in \llbracket 1..I \rrbracket, \forall j \in \llbracket 1..I-1 \rrbracket, \alpha_j^i = 0]\}$.

Proof (Lemma 2). Let's take α in \mathcal{R}^{I-1} and $\forall(i, j, m, p) \in (\llbracket 1..I \rrbracket)^4, i \neq j \neq m \neq p$:

$$\frac{\partial \mathcal{O}}{\partial \alpha_{i,j}} = \frac{\partial U_i}{\partial \alpha_{i,j}} + \frac{\partial U_j}{\partial \alpha_{i,j}} \Rightarrow \begin{cases} \frac{\partial^2 \mathcal{O}}{\partial \alpha_{i,j}^2} = h_i(C^i, \alpha^i) + s_i(\alpha^i) + h_j(C^j, \alpha^j) + s_j(\alpha^j) \\ \frac{\partial^2 \mathcal{O}}{\partial \alpha_{i,m} \partial \alpha_{i,j}} = s_i(\alpha^i) \\ \frac{\partial^2 \mathcal{O}}{\partial \alpha_{p,m} \partial \alpha_{i,j}} = 0 \end{cases}$$

Let's now prove that the Hessian matrix (\mathbf{H}) of \mathcal{O} is negative definite. We will, from now on consider only $-\mathbf{H}$ and prove that it is definite positive, under the assumption that $\nexists(i, j) \in (\llbracket 1..I \rrbracket)^2 / i \neq j, s_i(\alpha^i) = 0 \wedge s_j(\alpha^j) = 0$. We prove it by induction over I ($I > 1$). We denote \mathbf{H}_I the Hessian matrix for I , $\mathbf{H}_{I,(i,j)}$ its ij th-component, and $n_I = \binom{I}{2}$ the dimension of α and C .

Initialization: $I = 2 \Rightarrow n_2 = 1 \det(-\mathbf{H}_2) = -(h_1 + s_1 + h_2 + s_2) > 0$.

Inheritance: Assume for a given $I > 1$ that $-\mathbf{H}_I$ is positive definite.

$$\mathbf{H}_{I+1} = \begin{pmatrix} \mathbf{H}_I & R \\ R^T & Q \end{pmatrix}$$

with $\mathbf{H}_{I+1} \in \mathcal{M}_{n_{I+1}}(\mathbb{R}), \mathbf{H}_I \in \mathcal{M}_{n_I}(\mathbb{R}), Q \in \mathcal{M}_{(n_{I+1}-n_I)}(\mathbb{R}), R \in \mathcal{M}_{n_I \times (n_{I+1}-n_I)}(\mathbb{R}), \mathbf{H}$ and Q being both symmetric matrices by construction.

$$R = \begin{pmatrix} \frac{\partial^2 \mathcal{O}}{\partial \alpha_{1,2} \partial \alpha_{1,I+1}} & \cdots & \frac{\partial^2 \mathcal{O}}{\partial \alpha_{1,2} \partial \alpha_{I,I+1}} \\ \vdots & & \vdots \\ \frac{\partial^2 \mathcal{O}}{\partial \alpha_{I-1,I} \partial \alpha_{1,I+1}} & \cdots & \frac{\partial^2 \mathcal{O}}{\partial \alpha_{I-1,I} \partial \alpha_{I,I+1}} \end{pmatrix} \quad Q = \begin{pmatrix} \frac{\partial^2 \mathcal{O}}{\partial \alpha_{1,I+1}^2} & \cdots & \frac{\partial^2 \mathcal{O}}{\partial \alpha_{1,I+1} \partial \alpha_{I,I+1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{O}}{\partial \alpha_{I,I+1} \partial \alpha_{1,I+1}} & \cdots & \frac{\partial^2 \mathcal{O}}{\partial \alpha_{I,I+1}^2} \end{pmatrix}$$

To prove that $-\mathbf{H}_{I+1}$ is positive definite, we proceed by leading principal minors. As $-\mathbf{H}_I$ is assumed to be positive definite, we only need to consider the $n_{I+1} - n_I$ leading principal minors. We take $0 < i \leq n_{I+1} - n_I$ and denote $q_{i,j}$ the elements of Q , R_i the i th column of R and \mathbf{H}_{I+1,n_I+i} the matrix formed by the first $n_I + i$ rows and columns of \mathbf{H}_{I+1} . The i th leading minor (μ_i) that we have to consider can be written as:

$$\mu_i = |-\mathbf{H}_{I+1,n_I+i}| = \begin{vmatrix} -\mathbf{H}_{I+1,n_I+i-1} & -(R_i^T, q_{1,i}, \dots, q_{i-1,i})^T \\ -(R_i^T, q_{1,i}, \dots, q_{i-1,i}) & -q_{i,i} \end{vmatrix}$$

As R_i is the vector $(\frac{\partial^2 \mathcal{O}}{\partial \alpha_{1,2} \partial \alpha_{i,I+1}}, \dots, \frac{\partial^2 \mathcal{O}}{\partial \alpha_{I-1,I} \partial \alpha_{i,I+1}})^T$, which is composed of at least one s_i (the others elements being 0), and $(q_{1,i}, \dots, q_{i-1,i})$ being $(s_{I+1}, \dots, s_{I+1})$, we have that $(R_i^T, q_{1,i}, \dots, q_{i-1,i}) \neq 0_{\mathbb{R}^{n_I+i}}$ iff s_i (or s_{I+1} if $i > 1$) is non-null.

- if s_i (and s_{I+1} if $i > 1$) is null:

$$\mu_i = \det(-q_{i,i}) \det((-\mathbf{H}_{I+1,n_I+i-1})) > 0$$

- If s_i (or s_{I+1} if $i > 1$) is non-null, we have that $-(R_i^T, q_{1,i}, \dots, q_{i-1,i})$ is positive and $-q_{i,i}$ is strictly non-negative, and as $-\mathbf{H}_I$ is definite positive (and so is $(-\mathbf{H}_I)^{-1}$), we have that $\mu_i > 0$ by the block matrix determinant formula:

$$\begin{aligned} \mu_i &= \det(-q_{i,i}) \det \left((-\mathbf{H}_{I+1,n_I+i-1}) - \frac{1}{q_{i,i}} (-(R_i^T, q_{1,i}, \dots, q_{i-1,i}) (-(R_i^T, q_{1,i}, \dots, q_{i-1,i})^T)) \right) \\ \mu_i &> 0 \Leftrightarrow \det \left((-\mathbf{H}_{I+1,n_I+i-1}) - \frac{1}{q_{i,i}} (R_i^T, q_{1,i}, \dots, q_{i-1,i}) (R_i^T, q_{1,i}, \dots, q_{i-1,i})^T \right) > 0 \\ &\Leftrightarrow \det \left((-\mathbf{H}_{I+1,n_I+i-1}) - \frac{1}{q_{i,i}} (R_i^T, q_{1,i}, \dots, q_{i-1,i}) (R_i^T, q_{1,i}, \dots, q_{i-1,i})^T \right) \\ &\geq \underbrace{\det(-\mathbf{H}_{I+1,n_I+i-1}) - \frac{1}{q_{i,i}} \det((R_i^T, q_{1,i}, \dots, q_{i-1,i}) (R_i^T, q_{1,i}, \dots, q_{i-1,i})^T)}_{=0} > 0 \end{aligned}$$

Since $-\mathbf{H}_{I+1,n_I+i-1}$ was definite positive (if $i = 1, \mathbf{H}_{I+1,n_I+i-1}$ is \mathbf{H}_I), and as $\mu_i > 0$, $-\mathbf{H}_{I+1,n_I+i}$ is definite positive. We repeat this procedure until $i = n_{I+1} - n_I$ and get that $-\mathbf{H}_{I+1}$ is definite positive too.

If the assumption that $\nexists(i, j) \in ([1..I])^2, i \neq j, s_i(\alpha^i) = 0 \wedge s_j(\alpha^j) = 0$ does not hold, it is easy to notice that \mathbf{H}_I will have at least one column and row that is full of 0, thus $\det(\mathbf{H}_I) = 0$ in this case.

Hence, in the feasible set, \mathbf{H}_I is negative definite and as the feasible set is trivially connected we have that \mathcal{O} is strictly concave on the feasible set. \square

Theorem 6 (Unicity of the optimal graph). *For any $\mathcal{G}(\mathcal{S}, \alpha, C)$, the Social Optimization Problem has only one unique solution.*

Proof (Theorem 6). We can define the feasible set as $\{\alpha \in \mathcal{R}^{I-1}, [\alpha \in [0, 1]^{n_I}] \wedge [\nexists i \in [1..I], \forall j \in [1..I-1], \alpha_j^i = 0]\}$, and observe that it is convex. By Lemma 2, we have that the objective function (restricted to its first argument) is strictly concave on the feasible set. As constraints are qualified and convex, we have that the SOP has one unique solution. \square

Remark 3. We denote as "unconstrained solution" for a constrained problem a solution that corresponds to a critic point of the objective function.

Theorem 7 (Marginal utilities). *For a given individuals graph $\mathcal{G}(\mathcal{S}, \alpha, C)$, if α the solution of the Social Optimization Problem is unconstrained, the marginal utility of agent i with respect to $\alpha_{i,j}$ is equal to the reciprocal of the marginal utility of agent j with respect to $\alpha_{i,j}$.*

Proof (Theorem 7). For i and j different and in $[1..I]$, $\forall k \in \{i, j\}, \exists m_k \in [1..I-1], \alpha_{p_k}^k = \alpha_{i,j}$ by construction of α .

$$\frac{\partial \mathcal{O}}{\partial \alpha_{i,j}} = \sum_{i=1}^I \frac{\partial U_i}{\partial \alpha_{i,j}} = 0 \Leftrightarrow \frac{\partial U_i}{\partial \alpha_{p_i}^i} + \frac{\partial U_j}{\partial \alpha_{p_j}^j} = 0 \Leftrightarrow \frac{\partial U_i}{\partial \alpha_{i,j}} = -\frac{\partial U_j}{\partial \alpha_{i,j}}$$

\square

Theorem 8 (Invisible hand). *For any $\mathcal{G}(\mathcal{S}, \alpha, C)$ with \mathcal{C} the compatibility matrix defined as $\mathcal{C} = (c_{i,j})_{1 \leq i,j \leq N}$ (\mathcal{C}^i its i th-column), and $(\alpha^i)^*$ the solution vector of the Individual optimization problem (IOP) evaluated on \mathcal{C}^i , the compatibility list derived from \mathcal{C}_i , if*

1. $\exists \bar{c} \in \mathbb{R}^N$ s.t. $\forall i \in [1..N], \mathcal{C}^i \in \mathcal{S}(\bar{c})$, with $\mathcal{S}(\bar{c})$ the symmetric group of \bar{c}
2. there exists (H, S) two functions satisfying conditions of Definition 6 such that $\forall i \in [1..N], (\mathcal{S})_i = (P_i, U_i), \forall j \in [1..N-1], \frac{\partial U_i}{\partial \alpha_j^i} = H(c_j^i, \alpha_j^i) + S(\alpha^i)$

Then there exists $\hat{\alpha} \in \mathbb{R}^{n_N}$ such that $\forall i \in [1..N], (\alpha^i)^* \in \mathcal{S}(\hat{\alpha})$. Moreover, α^{**} the solution of the associated SOP, verifies:

$$\forall (i, j) \in [1..N]^2, \alpha_{i,j}^{**} = (\alpha^{\max\{i,j\}})_{\min\{i,j\}}^*$$

Proof (Theorem 8). Condition (2) implies that there exists some utility function \bar{U} such that $\forall i \in [1..N], U_i = \bar{U} + \epsilon_i$ with ϵ_i some real constant, i.e. all Individuals Optimization Problem (IOP) are the same. Condition (1) implies that each solution $(\alpha^i)^*$ (which exists by Theorem 5) must be a permutation of a vector $\hat{\alpha}$ of the same domain of α^i . As compatibilities are symmetric, even if the $(\alpha^i)^*$ are permutations, we have that the weight of the vertex that links individual i and j will be the same in both $(\alpha^i)^*$ and $(\alpha^j)^*$.

We prove this latter statement by contradiction. Let's take C that satisfies condition (1) and consider $((\alpha^k)^*, (\alpha^m)^*)$ the solution vectors of the IOP associated with two compatibility lists. For the sake of contradiction, we assume that $\alpha_m^k \neq \alpha_k^m$ but $((\alpha^k)^*, (\alpha^m)^*) \in (\mathcal{S}(\hat{\alpha}))^2$. We would have $\frac{\partial U_m}{\partial \alpha_k^m}(a) = H(c_k^m, a_k) + S(a), \frac{\partial U_k}{\partial \alpha_m^k}(a) = H(c_m^k, a_m) + S(a)$. By assumption, we have that $\alpha_m^k \neq \alpha_k^m$, implying that $\exists a \in \mathcal{S}(\hat{\alpha}), \frac{\partial U_m}{\partial \alpha_k^m}(a) \neq \frac{\partial U_k}{\partial \alpha_m^k}(a)$ but as $c_k^m = c_m^k$ and $\forall a \in \mathcal{S}(\hat{\alpha}), S(a) = \bar{s} \in \mathbb{R}$, the two functions are the same by (2), which is absurd.

Since we know by Theorem 6 that there exists a decentralized solution, and as the objective function of the SOP is $\mathcal{O} = \sum_{i \in [1..N]} U_i$, if α^{**} maximizes each U_i , it is the maximizer of \mathcal{O} . By definition U_i only depends on:

$$(\alpha^{**})^i = ((\alpha^{\max\{i,j\}})_{\min\{i,j\}}^*)_{1 \leq j \leq N, i \neq j}^i = ((\alpha^i)_1^*, \dots, (\alpha^i)_{i-1}^*, (\alpha^{i+1})_i^*, \dots, (\alpha^N)_i^*)$$

By the symmetry property that we have previously established, it can be rewritten as $(\alpha^{**})^i = ((\alpha^i)_1^*, \dots, (\alpha^i)_{i-1}^*, (\alpha^i)_{i+1}^*, \dots, (\alpha^i)_N^*)$ which is by assumption a maximizer of U_i . \square

Remark 4. Theorem 8 states that, if the two conditions are verified, all pairs of individuals (i, j) will desire the same $\alpha_{i,j}$ in the two individuals optimization problems of agents i and j (IOPs), and in the whole SOP.

Theorem 9. For any individuals set $\mathcal{I} \in (\mathcal{E}, \langle \cdot, \cdot \rangle) \times \mathcal{U}_N$, and weight list α of dimension $n_N = \binom{N}{2}$ satisfying:

- $\forall a \in \alpha, a \in [0, 1)$
- $\nexists i \in \llbracket 1..N \rrbracket, \forall j \in \llbracket 1..N-1 \rrbracket, \alpha_j^i = 0$
- $\forall M \leq I, \left(\sum_{i \leq M} H_i \right) (C, \alpha)$ is bijective in its first component.

There exists:

1. a unique compatibility list \bar{C} such that α is the unique unconstrained solution of the Social Optimization Problem for the individuals graph $\mathcal{G}(\mathcal{I}, \alpha, \bar{C})$
2. a set of compatibility lists \underline{C} such that $\forall C \in \underline{C}, \alpha$ is the unique solution of the Social Optimization Problem for the individuals graph $\mathcal{G}(\mathcal{I}, \alpha, C)$.

$$\underline{C} = \left\{ C \in (\mathbb{R}_+^*)^{n_I}, \quad c_{ij} \in \begin{cases} \{\bar{c}_{ij}\} & \text{if } \alpha_{ij} > 0 \\ (0, \bar{c}_{ij}] & \text{if } \alpha_{ij} = 0 \end{cases} \right\}$$

Proof (Theorem 9). Let's take $\mathcal{I} \in (\mathcal{E}, \langle \cdot, \cdot \rangle) \times \mathcal{U}_N$.

For point (1): We can construct a unique compatibility list (C) without any condition on α (for $i < j \leq N$).

$$(\nabla \mathcal{O})(\alpha) = 0 \Leftrightarrow \begin{cases} \vdots \\ \frac{\partial U_i}{\partial \alpha_{i,j}} + \frac{\partial U_j}{\partial \alpha_{i,j}} = 0 \\ \vdots \end{cases} \Leftrightarrow \begin{cases} \vdots \\ H_i(c_{i,j}, \alpha_{i,j}) + H_j(c_{i,j}, \alpha_{i,j}) + s_i(\alpha^i) + s_j(\alpha^j) = 0 \\ \vdots \end{cases}$$

As α is taken here as given (we are solving for C), each $c_{i,j}$ is determined by a single equation. As we have assumed that $\forall (i, j) \in (\llbracket 1..N \rrbracket)^2, H_i + H_j$ is bijective, we have that there exists a single C verifying $(\nabla \mathcal{O})(\alpha) = 0$ for α (and the individuals set) given.

From this compatibility list (C) and this set of individuals, we can form an individuals graph and write its associated SOP (Definition 10). By Theorem 6 this SOP has only one solution. By construction of C , α satisfies the first order condition of the optimization problem, and by assumption, α is in the feasible set. Hence, α is the unique solution of an SOP for the $\mathcal{G}(\mathcal{I}, \alpha, C)$ we have constructed.

For point (2): We can write the lagrangean of the Social Optimization problem (Definition 10):

$$\mathcal{L}(\alpha, C, \lambda, \mu) = -\mathcal{O}(\alpha, C) - \sum_{k=1}^{n_I} \lambda_k \alpha_k - \sum_{k=1}^{n_I} \mu_k (1 - \alpha_k)$$

We have already proven that the optimal solution of this problem is unique. We can write the first order conditions that characterize this solution:

$$\frac{\partial \mathcal{L}}{\partial \alpha_{i,j}} = -H_i(c_{i,j}, \alpha_{i,j}) - H_j(c_{i,j}, \alpha_{i,j}) - S(\alpha^i) - S(\alpha^j) - \lambda_{i,j} + \mu_{i,j} = 0$$

As $\forall (i, j) \in (\llbracket 1..N \rrbracket)^2, H_i + H_j$ is bijective, solving for $c_{i,j}$ is possible and has only one solution:

$$c_{ij} = (H_i + H_j)^{-1} (-\lambda_{ij} + \mu_{ij} - S(\alpha^i) - S(\alpha^j))$$

If no constraint is binding, that is if α is an unconstrained solution, $\lambda_{ij} = 0$ and $\mu_{ij} = 0$, and we have the same threshold as in point (1). If we have that α is a constrained solution ($\alpha_{i,j} = 0$), we have that $\mu_{ij} > 0$ and as $\forall (i, j) \in (\llbracket 1..N \rrbracket)^2, H_i + H_j$ is increasing, every value of $c_{i,j}$ that is under the threshold of (1) is possible. \square

Proposition 1. For any compatibility list C of dimension n_I , there exists \mathcal{I} an individual set of the space $\mathcal{I}^I = (\mathcal{E}, \langle \cdot, \cdot \rangle) \times \mathcal{U}_I$ with \mathcal{E} an euclidean space of dimension $K \in \mathbb{N}^*$. Moreover, $K \geq I$ is a sufficient condition for \mathcal{E} to exist.

Proof (Proposition 1). We first build a compatibility matrix $\tilde{C} \in \mathcal{M}(I, I)$ from the compatibility list C (and denoting P_i the characteristic vector in \mathcal{E} of the individual i):

$$(\tilde{C})_{i,j} = \begin{cases} c_{i,j} = \langle P_i, P_j \rangle & \text{if } i \neq j \\ \eta_i & \text{else} \end{cases}$$

First, \tilde{C} is symmetric by definition. Every real symmetric matrix has at least one positive definite submatrix. Let's denote \bar{L} the family of vectors P such that the submatrix $\tilde{C}_{\bar{L}}$ of their scalar product is positive definite. We take $\bar{L} \in \sup \{L : \tilde{C}_L \succ 0\}$ and denote $\bar{l} = \text{card}(\bar{L})$.

Second, we claim that if $\tilde{C}_{\bar{L}} \succ 0$ ($\tilde{C}_{\bar{L}}$ is positive definite), then the \bar{L} family of vectors P is linearly independent. We prove that claim by contradiction. Let's assume that we could have $\tilde{C}_{\bar{L}} \succ 0$ and \bar{L} not free.

$$\begin{aligned} \forall \mathbf{x} \in \mathbb{R}^{\bar{l}}, \mathbf{x}^t \tilde{C}_{\bar{L}} \mathbf{x} &= \sum_{i=1}^{\bar{l}} \sum_{j=1}^{\bar{l}} x_i (\tilde{C}_{\bar{L}})_{i,j} x_j \\ &= \sum_{i=1}^{\bar{l}} \eta_i + \sum_{i=1}^{\bar{l}} \sum_{j=1, i \neq j}^{\bar{l}} \langle x_i P_i, x_j P_j \rangle \end{aligned}$$

If we define $\eta_i = \langle P_i, P_i \rangle$, we have $\mathbf{x}^t \tilde{C}_{\bar{L}} \mathbf{x} = \left\| \sum_{i=1}^{\bar{l}} x_i P_i \right\|^2$.

$$\tilde{C}_{\bar{L}} \succ 0 \Leftrightarrow \forall \mathbf{x} \in \mathbb{R}^{\bar{l}}, \mathbf{x}^t \tilde{C}_{\bar{L}} \mathbf{x} > 0 \Leftrightarrow \forall \mathbf{x} \in \mathbb{R}^{\bar{l}}, \left\| \sum_{i=1}^{\bar{l}} x_i P_i \right\|^2 > 0$$

But if \bar{L} is linearly dependant, there exists \mathbf{x} such that $\left\| \sum_{i=1}^{\bar{l}} x_i P_i \right\|^2 = 0$ breaking the positive definiteness condition.

Finally, as every vector of \bar{L} is a vector of \mathbb{R}^K , we need K to be at least $\bar{l} = \text{card}(\bar{L})$. By definition, we have $K \geq I \geq \bar{l}$. \square

Definition 11 (Pairwise stability, from (Jackson 1996)). *A graph $\mathcal{G}(\mathcal{J}, \alpha, C)$ is pairwise stable, iff*

1. *for all (i, j) such that $\alpha_{i,j}$ is a non-null component of α , we have*

$$\forall k \in \{i, j\}, \quad U_k(\alpha, C) \geq U_k(\alpha', C)|_{\alpha_{i,j}=0}$$

2. *for all (i, j) such that $\alpha_{i,j}$ is null component of α , we have*

$$\forall a \in (0, 1), \forall k \in \{i, j\}, \quad U_k(\alpha, C) \geq U_k(\alpha, C)|_{\alpha_{i,j}=a}$$

Proposition 2. *Every graph $\mathcal{G}(\mathcal{J}, \alpha, C)$ that is solution of the Social Optimization Problem (Definition 10) is pairwise stable.*

Proof (Proposition 2). We prove both results (1 and 2) by contradiction. Assume that there could exist an α' in the feasible set such that $\exists (i, j) \in \llbracket 1..N \rrbracket^2, \forall k \in \{i, j\}, U_k(\alpha', C) \geq U_k(\alpha, C)$ and $\forall k \in \llbracket 1..N \rrbracket^2 \setminus \{(i, j), (j, i)\}, \alpha_k = \alpha'_k$. By construction of utility functions, we have that $\forall k \in \llbracket 1..N \rrbracket \setminus \{i, j\}, U_k(\alpha, C) = U_k(\alpha', C)$. Hence, we have $\sum_{k \leq N} U_k(\alpha', C) \geq \sum_{k \leq N} U_k(\alpha, C) \Rightarrow \mathcal{O}(\alpha', C) \geq \mathcal{O}(\alpha, C)$. As α is the solution of the SOP, and as the solution is unique, this inequality is absurd. \square

3. Alky utility function

Definition 12 (ALKY utility function). *We consider the set \mathcal{U}_I^{ALKY} of applications $U : [0, 1]^{I-1} \times (\mathbb{R}_+^*)^{I-1} \rightarrow \mathbb{R}$:*

$$\mathcal{U}_I^{ALKY} = \left\{ \kappa > 1, \gamma > 1, \delta > 1, U : (\alpha, c) \mapsto \sum_{j=1}^{I-1} \left[\kappa \alpha_j c_j - \frac{\alpha_j^\gamma}{1 - \alpha_j^\gamma} \right] - \left(\sum_{j=1}^{I-1} \alpha_j \right)^\delta \right\}$$

Remark 5. Analytically, this utility function is made of a three parts: a positive one which weighs affinity $(c_{i,j})_{i,j \leq N}$ by the relationship intensity $(\alpha_{i,j})_{j \leq N}$ and two costs, that convexly penalize intensity, and create substitution effects between relationships.

γ is the convexity parameter of relationships' direct cost (and prevents corner solutions). δ enable the substitutability effects between relationships. For $\delta > 1$ relations are substitutable: an agent prefers reducing the emphasis on relationships with lower affinity in order to increase the weight on his most successful matches.

Theorem 10 (ALKY Utility). \mathcal{U}_I^{ALKY} is a subset of \mathcal{U}_I , that is all ALKY functions are utility functions.

Proof (Theorem 10). Let's take $U \in \mathcal{U}_I^{ALKY}$, $i \in \llbracket 1..I-1 \rrbracket$ and $j \in \llbracket 1..I-1 \rrbracket, j \neq i$.

First, (1) and (2) are trivially met $\forall C \in \mathbb{R}^{n_I}, U(0, C) = 0$ and $\nabla U(0) > 0$ iff C is a strictly positive vector.

$$\begin{aligned} \frac{\partial U}{\partial \alpha_i} &= \underbrace{\kappa c_i - \frac{\gamma \alpha_i^{\gamma-1}}{(1 - \alpha_i^\gamma)^2}}_{H(c_i, \alpha_i)} \underbrace{- \delta \left(\sum_{k=1}^{I-1} \alpha_k \right)^{\delta-1}}_{S(\alpha)} \\ \frac{\partial^2 U}{\partial \alpha_i \partial \alpha_j} &= \underbrace{-\delta(\delta-1) \left(\sum_{k=1}^{I-1} \alpha_k \right)^{\delta-2}}_{s(\alpha)} < 0 \\ \frac{\partial^2 U}{\partial \alpha_i^2} &= \underbrace{\frac{\gamma \alpha_i^{\gamma-2}}{(1 - \alpha_i^\gamma)^3} \left[(1 - \gamma) - (1 + \gamma) \alpha_i^\gamma + 2\gamma \alpha_i^{2\gamma} \right]}_{h(c_i, \alpha_i)} \underbrace{- \delta(\delta-1) \left(\sum_{k=1}^{I-1} \alpha_k \right)^{\delta-2}}_{s(\alpha)} \end{aligned}$$

If $s(\alpha)$ is trivially strictly negative for all non-zero vector, we need to consider the sign of $h(c_i, \alpha_i)$. As $\gamma > 1, \delta > 1$, the only remaining term of undetermined sign is: $(1 - \gamma) - (1 + \gamma) \alpha_i^\gamma + 2\gamma \alpha_i^{2\gamma}$. We switch variables by defining $X = \alpha_i^\gamma$ and get the following second order polynomial: $(1 - \gamma) - (1 + \gamma)X + 2\gamma X^2$ evaluated on $[0, 1)$. It is easy to observe that the polynomial is convex, and is negative on $[0, 1)$. Hence, $h(c_i, \alpha_i) < 0$. □

Remark 6. Any individuals set that is only composed of ALKY utility functions satisfies the conditions of Theorem 9: as ALKY functions are linear in C , the sum of their partial derivative with respect to α is trivially bijective.

Appendix B: Detailed algorithms for the agent-based model

Algorithms 3 and 4 are provided as they are closer to the implementation in C++. $\lambda = 0.1$ is the learning rate, $\Omega = 0.1$ the max step size, $\omega = 0.05$ is the minimum weight (above which the link cannot exist) and $\varpi = 0.0001$ the minimum accepted gradient value. In the C++ implementation, the weights are computed using sigmoid, and the gradient move is made on the argument of sigmoid and not directly on the weight.

Algorithm 3 Individual i select action protocol (Detailed)

```

function TAKEACTION( $i$ )
   $\mathcal{S}^i \leftarrow \{j : d(j, i) \leq \delta\}$                                 ▷ Get individuals in the scope of  $i$ 
   $C_S^i \leftarrow \{c_{i,j} : j \in \mathcal{S}^i\}$                                 ▷ Get the set of compatibilities
   $\alpha_S^i \leftarrow \{\alpha_{i,j} : j \in \mathcal{S}^i\}$                                 ▷ Get the set of weights
  for  $j \in \mathcal{S}^i$  do
    if  $j \in M$  then                                              ▷ Check the memory set
       $\nabla U_{(j)}^i(\alpha_S^i, C_S^i) \leftarrow 0$ 
    else if  $\nabla U_{(j)}^i(\alpha_S^i, C_S^i) \leq \varpi$  then
       $\nabla U_{(j)}^i(\alpha_S^i, C_S^i) \leftarrow 0$ 
    end if
  end for
  if  $\text{card}(M) > 3$  then
    FORGETAMEMORY                                              ▷ Delete one memory
  end if
   $k \leftarrow \text{softargmax}_j |\nabla U^i(\alpha_S^i, C_S^i)|$ 
  if  $\nabla U_{(k)}^i > 0$  then
     $M \leftarrow M \cup k$                                           ▷ Remember the action
    return REQUEST( $k, i, \alpha_{i,k} + \lambda \nabla U_{(k)}^i$ )                ▷ Ask  $k$  to increase weight
  else if  $\nabla U_{(k)}^i < 0$  then
     $M \leftarrow M \cup k$                                           ▷ Remember the action
     $\alpha_{i,k} \leftarrow \alpha_{i,k} + \lambda \nabla U_{(k)}^i$                     ▷ Direct decrease in friendship with  $N^k$ 
    return true
  else
    return false                                              ▷ No action has been taken
  end if
end function

```

Algorithm 4 Individual k response protocol to i (request function, detailed)

```

function REQUEST( $k, i, a$ )
   $\mathcal{S}^k \leftarrow \{j : d(j, k) \leq \delta\} \cup \{i\}$                                 ▷ Get individuals in the scope
   $C_S^k \leftarrow \{c_{k,j} : j \in \mathcal{S}^k\}$                                 ▷ Get the set of compatibilities
   $\alpha_S^k \leftarrow \{\alpha_{k,j} : j \in \mathcal{S}^k\}$                                 ▷ Get the set of weights
  if  $\nabla U_{(i)}^k(\alpha_S^k, C_S^k) > 0 \wedge \min(a, \alpha_{i,k} + \lambda \nabla U_{(i)}^k) > \omega$  then
    if  $\alpha_{i,k} > 0$  then
       $\alpha_{i,k} \leftarrow \min(\min(a, \alpha_{i,k} + \lambda \nabla U_{(i)}^k), \alpha_{i,k} + \Omega)$ 
    else
       $\alpha_{i,k} \leftarrow \Omega$ 
    end if
    return true
  else if  $\nabla U_{(i)}^k(\alpha_S^k, C_S^k) < 0$  then                                ▷ If he wants to reduce the weight
     $\alpha_{i,k} \leftarrow \max(\max(\alpha_{i,k} + \lambda \nabla U_{(i)}^k, \alpha_{i,k} - \Omega), 0)$ 
    return true
  end if
  return false
end function

```

Algorithm 5 Simulation body (detailed)

```

 $t \leftarrow 1$                                 ▷ Initialize the time (round counter)
 $v \leftarrow 0$                                 ▷ Initialize an inactive rounds counter
for  $t \leq \text{MAX-SIMULATION-TIME} \wedge v \leq 50 + 3$  do
   $v_t \leftarrow 0$ 
  for  $i \in o_t \in \mathcal{S}(\llbracket 1..N \rrbracket)$  do                                ▷ For each individuals in a random order
    if  $\text{TAKEACTION}(i) = \text{false}$  then                                ▷ If no action has been taken
       $v_t \leftarrow v_t + 1$ 
    end if
  end for
  if  $v_t = I$  then
     $v \leftarrow v + 1$ 
  end if
  if  $t > \text{MAX-SIMULATION-TIME} \times 10\%$  then
    for  $\alpha \in \{\alpha_{i,j} < \omega : i \leq I, j \leq I\}$  do
       $\alpha \leftarrow 0$                                 ▷ Removal of the smallest edges
    end for
  end if
end for                                ▷ End of the simulation process

```
