Generate country-scale networks of interaction from scattered statistics

Samuel Thiriot[1,2] ⋆ and Jean-Daniel Kant[1]

[1] Computer Science Laboratory - University of Paris 6 (LIP6), France
[2] France Télécom R&D - Orange Labs
Samuel.Thiriot@lip6.fr, Jean-Daniel.Kant@lip6.fr

Abstract. It is common to define the structure of interactions among a population of agents using the social network metaphor. Most agent-based models were shown to be highly sensitive to that network, so the relevance of simulation results depends directly on the descriptive power of that network. When studying social dynamics in large populations, this network cannot be collected, and is rather generated by algorithms which aim to fit general properties of social networks. However, more precise data is available at a country scale in the form of socio-demographic studies, census or sociological studies. These "scattered statistics" provide rich information, especially on agents' attributes, similar properties of tied agents and affiliations.
In this paper, we propose a *generic methodology* to bring together these scattered statistics with Bayesian networks. We explain how to *generate a population of heterogeneous agents*, and how to create links by using both scattered statistics and knowledge on social selection processes. The methodology is illustrated by generating an interaction network for rural Kenya which includes family structure, colleagues and friendship. That network is constrained by data available from statistics and field studies.

1 Context & problematic

1.1 Problematic

The principle of agent-based models is to reproduce collective dynamics from local interactions. So in any agent-based simulation, the modeler requires a descriptive model of interactions in a population. As these relationships are relatively stable [1], it became common to represent them using the *social network metaphor*: the structure of interactions is represented by a graph $G(\mathcal{A}, \mathcal{L})$, with $\mathcal{A}$ the population of agents and $\mathcal{L}$ the links between these agents. This structure was shown to have a dramatic influence on the dynamics of various agent-based models. The direct consequence is that *the descriptive power of the structure determines the relevance of simulation results*. To ensure that the generated relationship network is descriptive, it should be studied as a modeling problematic:

⋆ This work was partially funded by grant CIFRE 993/2005 from the French National Association for Research and Technology (ANRT).

the structure of relationships should be a simplification of social interactions, and comply with knowledge on the modeled population.

While the interaction network can be collected by interview when the population is small, such a data collection becomes intractable for larger populations. Hence, a lot of models deal with *country-scale populations*, including models of opinion dynamics, virus propagation or information dynamics. In our case, we are interested in modeling diffusion of innovations [2], and propagation of information about these innovations [3]. In this field, the lack of descriptive model of interactions was pointed out as one fundamental limitation of models [4]. It is common to use *network generators* to describe such a large population. A network generator is an algorithm which, given several parameters, generates networks compliant with one or more properties observed in real networks.

Ideally, a network generator should satisfy the following requirements (noted R). **(R1) Generate models of large populations**, in order to improve the descriptive power of agent-based models for large-scale simulations. **(R2) Represent the different kinds of relationships linking two agents**, because interactions do not occur in the same way across different relationships. For instance, finding a job was shown to be more efficient when activating so-called "weak ties" (for instance far family) [5]. **(R3) Detail attributes of agents** in the network of interactions. That is justified by three main reasons. First, *the attributes of individuals influence their judgment and decision-making* (*e.g.* in diffusion of innovations [4]), so they should be available in the model. Secondly, *attributes influence the frequency or nature of interaction*: spatial distance reduces frequency of exchanges, differences in ethnicity and interests decrease the normative influence, etc. Thirdly, it was shown (as explained in 1.2) that *individual characteristics determine the choice of acquaintances* of an agent, therefore agents' traits should be taken into account during network generation.

1.2 Key findings for social networks

Decades of research in social networks highlighted several key findings which today are widely accepted. A stream of research explored *social selection processes* [6] in order to understand *how agents create ties*. It appeared that individuals exhibit a strong tendency to create relationships with people who share similar characteristics (*homophily*) [7]. Two individuals sharing a common *affiliation* (event, project or workplace) also have more chances to bond and interact frequently [1]. *Transitivity* implies that two individuals sharing a common acquaintance are more probably connected together; in fact they have more chances to meet and to create a tie because of a common friend. These observations are no longer questioned, so **(R4) a relevant network of interactions should comply with processes of social selection**.

The stream of social network analysis also described *statistical properties shared by social networks*, including a surprisingly short average distance between individuals, a high clustering rate (there are groups or communities in which individuals are strongly interconnected), and a low density [1]. A power-law distribution of degrees was also observed in various datasets (most individuals have few acquaintances, while few have a high degree of connectivity). It was explained by the so-called preferential attachment principle, which states that new individuals in a network will connect more probably with nodes which already have a high degree of connectivity.

Beyond these general properties of social networks, each national institute of statistics publishes detailed data for its country. Statistics describe *who the individuals are* by quantifying characteristics like gender, age, ethnicity, socioeconomic class, incomes, marital status, etc. They also study *what people do*: being employed or not, type of activity, participation in associative life, sport, etc. These activities can often be interpreted as *affiliations*, with detailed information on agents which are part of the institution (common characteristics of workers like educational level, socioeconomic class or geographical location) and on the affiliations themselves (size, location). More qualitative knowledge also exists on the structure of families, as well as statistics about the number of children or household composition. When this kind of data is not collected at a large scale, it is still available from field studies focused on more precise phenomena. As an illustration for this paper, we chose to model social relationships in rural Kenya, for which we have no information of our own. Demographic statistics [8], sociological studies on the structure of families [9] and field studies on diffusion of contraceptive use (*e.g.* [10,11,12]) constitute many sources of information. Surprisingly, no network generator uses these scattered statistics. However, they constitute an appreciable part of knowledge on the structure of interactions. We claim that these **(R5) scattered statistics should be taken into account while generating a network of interactions**.

1.3 Existing models

The generators mostly used for agent-based models are small-world networks and scale-free graphs [13]. The first generates networks which are highly clustered with a short average path length, while the second implements the preferential attachment principle. These models, proposed by physicists, generate highly stylized networks in which social relationships between heterogeneous agents become links between nodes. They neither comply with knowledge of social selection processes, nor rely on statistics available for a given population. To summarize, they do not satisfy our requirements R2, R3, R4 or R5.

In the context of social network analysis [1], a lot of models were proposed (see [6] for a synthetic picture). Existence of a link between two agents $a_1, a_2 \in \mathcal{A}$ is considered to be a random variable L^{a_1,a_2} which takes the value 1 if a link exists, and value 0 if not. Random graphs with attributes generate links given a vector of agents' attributes Att: $p(L^{a_1,a_2} = 1|Att(a_1), Att(a_2))$. If one uses only this constraint, a tie between two agents is independent of any other tie with other agents, in contradiction with transitivity evidence. This assumption was removed by markov random graphs [14] by allowing two links to be dependent if they have a node in common. In this case links are noted as the conditional probability: $p(L^{a_1,a_2} = 1|L^{a_1,a_3} = 1, L^{a_3,a_2} = 1)$, with $a_1, a_2, a_3 \in \mathcal{A}$. Recent

extensions of these models [6] take into account both links created given agents' attributes and transitive links; to date, they remain limited to one or two attributes at most [6].
This formalism is powerful enough to describe homophily and transitivity. Its relevance was proved by fitting data collected from small groups. It was also shown [6] that affiliations or degree of an agent may be considered to be attributes, so the formalism also enables the generation of graphs with power-law degree distribution and affiliations. In short, they fulfill R3 and R4. However, they include special parameters which require ad hoc collection of data, so their application remains limited to small groups (in opposition with R1). Moreover, in these small groups, it was never necessary to distinguish different kinds of relationships, contrary to R2.

1.4 Approach

We propose to use scattered statistics (R5) available for a given country (R1) to generate more representative networks of interactions. The generated relationship network $G(\mathcal{A}, \mathcal{L}, Att, \mathcal{T})$ will detail agents' attributes Att in the population (R3), which are taken into account during the network generation. The network will include several kinds of relationships $\mathcal{T}$ (R2), so the user of this network may infer the interaction network given the kind of relationship (see 4.1). As the model is intended to bring together several sources of information on a population in order to parameter the generator, we describe a methodology (see 2) to intuitively formalize knowledge on agents' attributes and links using Bayesian networks. The formalism, inspired by markov random graphs, enables the representation of the key processes of social selection (R4). Then (in section 3), we explain how a population of heterogeneous agents can be generated and how the relationship network is created. Insights into the minimum size of population, detection of statistical discrepancies, and into statistical properties of generated networks are provided in section 4.

2 Methodology

2.1 Choice of agents' attributes and link types

Step 1 The modeler should firstly **define the types of social links** $\mathcal{T}$ that should be represented in the relationship network. $\mathcal{T}$ enumerates links potentially leading to different interactions in the model, or those created by different processes. As proposed previously in markov random graphs [14], some kinds of relationships can be generated given agents' attributes $\mathcal{T}^{Att}$, while others are created by transitivity $\mathcal{T}^{trans}$. In the example of Kenya, we choose to represent links leading to interaction about contraceptive use. Field studies indicate that spouses discuss that topic, parental advice has a normative influence, and that women retrieve information from friends, siblings and colleagues [11]. We define links created given agents' attributes $\mathcal{T}^{Att} = \{$*spouses, motherOf, colleagues, friends*$\}$. Links involving more than two agents depend on links already created, and are formalized by transitivity: $\mathcal{T}^{trans} = \{$*fatherOf, siblings*$\}$.

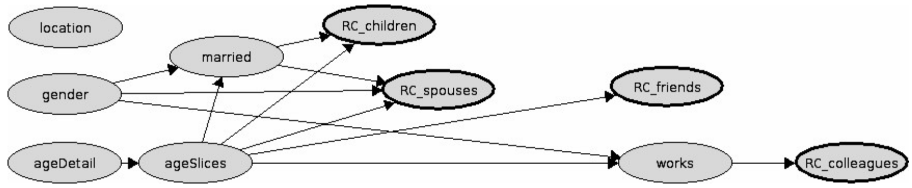

Fig. 1. The agent BN used to describe interdependencies between Kenyan socio-demographic attributes. Nodes in bold are the number of links to create for each link type.

Step 2 Next, the modeler has to **select agents' attributes** *Att* which are known - or supposed - to influence the probability of a link to be created. Of course, that selection is conditioned by data availability, and takes into account the purpose of the agent-based model. Typically agents' attributes contain *socio-demographic characteristics* (age, gender, socioeconomic class, ethnicity, etc.) and *places where the agents have frequent interactions* (going to school, frequenting a workplace, etc.). We assume the number of links per agent for each kind of relationship to be an attribute: $\forall t \in \mathcal{T}$, $RC_t^a \in Att(a)$. This choice of including the number of links as an attribute could seem counter-intuitive, because it was often considered to be an independent density parameter [6]. This choice is justified by the following reasons: (i) the number of links per agent is available from statistics, and varies across the kinds of relationships. (ii) *The number of links is strongly correlated to other agents' attributes*; for instance the number of children of a married woman depends on her age. (iii) *The number of links is often considered to be an explanatory variable* for the individual decision-making process, therefore it should be implemented as an attribute. For example, adoption of contraceptive by a woman increases with the number of its children [10]. In the example of Kenya, *marital status*, *age* and *gender* attributes are required for nearly all kinds of links. As field studies indicate that most discussions take place during daily activities [12], we added the variable *work* to represent them, with possible values "collect water" or "work in market". *Spatial location* is also required because spouses always live in the same place; in the same way, young children always grow up near their mother.

2.2 Formalization based on Bayesian networks

Step 3: represent agents' attributes using Bayesian networks. Attributes of individuals in a real population are strongly interdependent: marital status depends on age and gender, socioeconomic class is highly correlated with location and education level, etc. Generate an heterogeneous population of agents requires a relevant formalism for these complex interdependences. That formalism should be generic enough to represent any kind of data. Most of the time, data available for a population is presented as statistics linking one attribute to another. For instance, the number of children per woman is provided given marital status and age [8, p. 57]. This kind of statistic can be translated, without loss of generality, to conditional probabilities, like: $p\left(RC^{a}_{motherOf} = \{0...10\} \mid \text{age}(a), \text{ married}(a)\right)$. In this viewpoint, *attributes of agents are considered to be random variables*. We propose to use a Bayesian network [15] (BN), named *agent BN* in this methodology, in order to formalize these interdependencies. Each agent attribute in *Att* is represented by a variable in the BN. The domain of a variable defines the values the attribute can take. For instance in graph 1, the variable *gender* has for domain $D^{gender} = \{male, female\}$, $D^{married} = \{yes, no\}$, and $D^{RC\_motherOf} = \{0...10\}$. Root variables define initial probabilities. In Fig 1, the variable *ageDetail* defines the probability for an individual picked randomly from the population to be aged 0,1,2 up to 100 years; this probability is available from the age pyramid of the target population. A directed link between two variables $V_1 \rightarrow V_2$ means that V_2 probabilities can be calculated using the parents, and only the parents. V_2 embodies a conditional probability table representing the probability to take each value D^{V_2} given all the possible values in the domains of its parents (here, V_1). No link means that variables are assumed independent. It does not mean that variables are independent in reality, but rather represent our lack of knowledge (or our willingness to simplify that knowledge) of the dependencies.

In our application to Kenya, probabilities in the agent BN depicted in Fig. 1 come from the US Census Bureau, from the Kenyan demographic and health survey [8], and from field studies (e.g. [10,11,12]). Note that we used *convenience variables* to simplify formalization of data: in agent BN 1, the variable *ageSlices* simplifies the detailed age to 5-year slices, which are often used in published statistics. Another benefit of BN is the ability to *highlight evident discrepancies in data*. For instance, a social scientist will immediately notice in (Fig. 1) the absence of a link between gender and age, while the age pyramid in most of countries shows significant differences between genders (indeed, Kenya is a particular case of symmetrical age pyramid).

Step 4: represent links probability using Bayesian networks. Links created given attributes $\mathcal{T}^{Att}$ are defined by $p(L_t^{a_1,a_2} = 1 | Att(a_1), Att(a_2))$. They can be used to represent a large range of phenomena, including *homophily*, *affiliation*, *preferential attachment*, or spatialization. As this probability is conditional, it can also be represented by a *matching BN*. An example of such a BN describing relationship *spouses* in Kenya is depicted in Fig. 2. In this matching BN, one can recognize two instances of the agent BN (on the left) representing attributes of two different agents of the population. On the right, a special node with domain $\{yes, no\}$ define whether a link can be created between these agents. Nodes in bold define constraints on linkage. In this BN, we arbitrarily define that agent 1 (top) is a male and agent 2 (bottom) a female (marriage in Kenya is heterosexual). Node *ageWife* projects the probable age of the wife of the man a_1 (on average 10 years younger), and variable *rightAge* ensures that the wife a_2 complies with that age (using an identity probability table). The node *sameLocation* only takes the value *yes* if both agents live in the same location. The final node

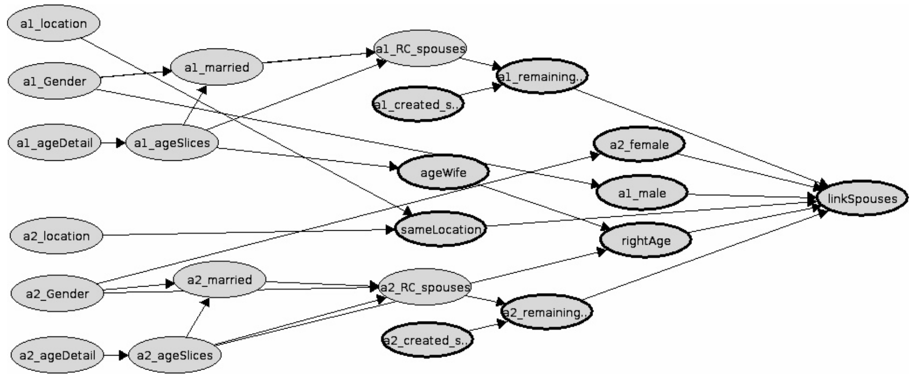

Fig. 2. Matching Bayesian Network for link type *spouses*. On the left, the agent BN for agents 1 and 2.

"linkSpouses", which determines whether two agents can be linked together, takes the value "yes" only if all of its parents have for values "yes" (AND operator). The nodes *a1_created_spouses* and *a1_remaining_spouses* ensure that we will only create as many links of type t as required by $RC_t^{a_1}$ and $RC_t^{a_2}$ and no more. Thus a wife will only be linked with one husband. Other links in $\mathcal{T}^{Att}$ are defined in the same way. Friends are probably the same age and probably live in the same town. Mothers are linked to children whose age is compliant with their own age, and they always live in the same location when these children are young. Colleagues are defined as agents sharing the same activity in the same location.

Links created by transitivity $\mathcal{T}^{trans}$ are also random variables. They are noted: $p(L_{t_1}^{a_1,a_2} = 1 | L_{t_2}^{a_1,a_3} = 1, L_{t_3}^{a_3,a_2} = 1)$, with $a_1, a_2, a_3 \in \mathcal{A}$, $t_1 \in \mathcal{T}^{trans}$, $t_2, t_3 \in \mathcal{T}$. This formalism is quite intuitive and will not be detailed further here. In our example, the link "fatherOf" is defined by transitivity with $p(L_{fatherOf}^{a_1,a_2} = 1 | L_{motherOf}^{a_1,a_3} = 1, L_{spouses}^{a_3,a_2} = 1) = 1$ (only transitivity enables the description of father-children links; it couldn't be described as a matching BN, because a man's children have to be the same as those of his wives). In the same way, siblings are created by transitivity across mother and father links. With a lower probability, friendships links are created by transitivity between friends.

3 Generation of the graph

3.1 Generation of a heterogeneous population

All the variables *Att* in the agent BN will become agents' attributes with the same domain. For each agent $a \in \mathcal{A}$ to be created, we generate a *prototype agent*. The process to generate a prototype simply consists of using the agent BN in a generative way: for each variable $V \in Att$ of the agent BN (in the ordinal order, so root variables are processed first), a value $V = v$ is selected randomly

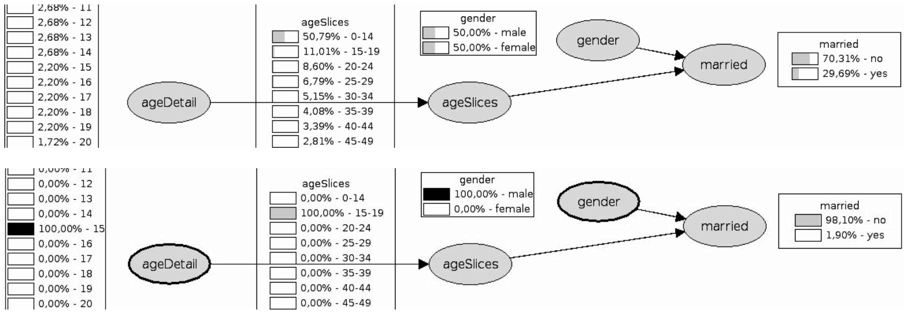

Fig. 3. Example of evidence propagation when the Bayesian network is used to generate agents' attributes. Here monitors (boxes in the figure) display the probabilities for each variable to take each value (note that some of these monitors are truncated). *(top)* probabilities with no evidence *(bottom)* probabilities when evidence is set.

in the domain of V, given probabilities $p(V = v|parents(V))$ defined in the BN. When value $V = v$ has been chosen, a corresponding piece of evidence $p(V = v) = 1$ is put in the BN. Evidence, in the theory of BN, represents a known information. Putting evidence in the BN allows the computation of child variables given the values of attributes already selected. For instance in Fig. 3, before any piece of evidence *(top)*, the probability for someone randomly picked from the population being married is 29.69%. When attributes *ageDetail* and *gender* have been randomly selected respectively to 15 and *male*, and used as evidence, the posterior probability for the current agent to be married falls to 1.90%. When all the agents are generated in this way, the statistical distribution of their attributes complies with the distribution described by the BN.

3.2 Creation of links

Now that all agents have been created in the population $\mathcal{A}$, each agent having its attributes $Att(a)$ defined, we have to link them by using the matching BN. For each kind of relationship $t \in \mathcal{T}^{Att}$, we constraint the matching BN for t by providing *evidence for link creation*: as our aim is to link together agents with link t, we set evidence on variable $p(link\_spouses = yes) = 1$. Given this evidence, probabilities for attributes of a_1 and a_2 are updated, and some probabilities in attributes' domains fall to zero. For instance, in the case of "spouses" link, the probability of agents 1 and 2 being younger than 15 years falls to zero; they also cannot have "married=no". In other words, probabilities in the matching BN given evidence of link creation designate two sets of *candidates for linking* $\mathcal{C}_1^t$ and $\mathcal{C}_2^t$. The matching process will remain limited to these sets. Then, we iterate across candidates $\mathcal{C}_1^t$ and randomly select as many acquaintances among $\mathcal{C}_2^t$ as required by $RC_t^{a_1}$. For each agent $a_1 \in \mathcal{C}_1^t$, we load its attributes and use them as pieces of evidence in the BN. After a run of the inference engine, the probabilities for agent 2 define a restricted set $\mathcal{C}_2^t|Att(a_1) \subset \mathcal{C}_2^t$ of candidates for linkage *given agent 1 attributes*. In our application to Kenya, for link type

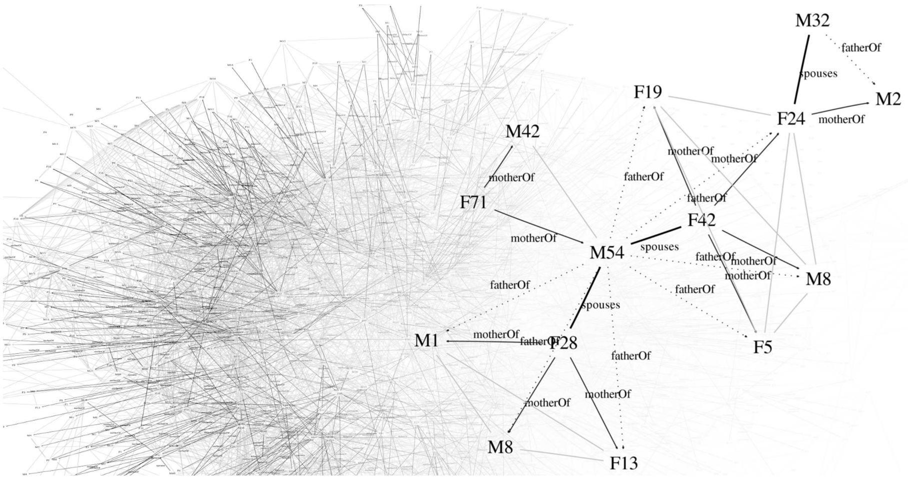

Fig. 4. Generated relationship network *(left)* Zoom in one agent *(right)*

spouses, C_1 is the set of husbands and C_2 the set of wives. When one chooses an agent a_1, given the constraints on matching, the set $\mathcal{C}_2|Att(a_1)$ limits $\mathcal{C}_2$ to wives who live in the same location as a_1. The selection of a candidate is done by generating a prototype agent as explained before; if that prototype cannot be found, a fallback solution consists of randomly picking one agent in $\mathcal{C}_2^t|Att(a_1)$ (note that the fallback solution can bias statistical distribution in the population; in our example, it is possible to link a husband with an older wife). When no fallback solution can be found, this is when $\mathcal{C}_2|Att(a_1) = \{\varnothing\}$, agent a_1 remains orphan, but will never be tied with an incompatible agent (in our example, no man will be said to be married with a too young woman nor with a woman having attribute "married=no"). These errors will be studied in section 4.2.

After having processed all link types defined by matching BN, transitive links are created using the probabilities $p(L_{t_1}^{a_1,a_2}|L_{t_2}^{a_1,a_3}, L_{t_3}^{a_3,a_2})$ formalized in step 2.

4 Generated network

4.1 Usage for social simulation

The resulting graph $G(\mathcal{A}, \mathcal{L}, Att, \mathcal{T})$ includes links $\mathcal{L}$ of different kinds ($\mathcal{L} = \bigcup_{t\in\mathcal{T}} \mathcal{L}^t$), and provides the values of agent attributes Att for any agent in the population $\mathcal{A}$. The structure of relationships described by the generated network obviously depends on agent BN and matching BN provided by the modeler as parameters. In our application to Kenya, the population covers the whole age pyramid, and describes attributes depicted in Fig. 1. Moreover, as shown in Fig. 4, each agent is positioned in its family environment; agent M54 (for Male, 54 years) is married with two wives F28 and F42, and has 7 children, including one daughter F24 who herself is married and a mother. He is also tied with his

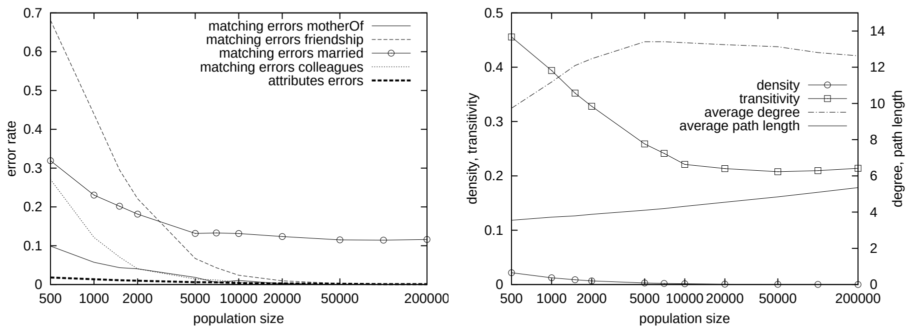

Fig. 5. Error rate given population size *(left)*. Statistical properties of the generated relationship network *(right)*

own mother F71 and brother M42, but not with his father - probably because he is not in the age pyramid (no longer alive). He his also tied with *colleagues* and *friends* (not represented in the figure to improve readibility). That structure is described at the scale of 50,000 agents depicted in Fig 4 *(left)*.
To use this relationships' network for simulation, the modeler may simply define probability to interact given the kind of relationship: $\forall t \in \mathcal{T}$, $p^t_{interact}$, so $p_{interact}(a_1, a_2 | L_t^{a_1,a_2}) = p^t_{interact}$. He/she may also choose a finer granularity by defining the probability of interaction given attributes Att, for instance to represent the fact that spatial distance decreases the probability of interaction. In this illustration, we focus on interactions about contraceptive use [11]. In our case, no interaction occurs across links between young children and their parents. As the topic of contraceptive use is sensitive in Kenya, probabilities of discussion between spouses are low, as between a mother and her own parents. In fact, women who are still fertile and are concerned by the topic mainly discuss it with their female friends, and often with their brothers-in-law (link sibling). The resulting network of interactions is a network in which links are weighted by probabilities; it is by far sparser than the network of relationships.

4.2 Errors and statistical properties

While BN describe a theoretical population with continuous probabilities, we generate a discrete population and link agents only when a suitable candidate exists. This limitation necessarily leads to a bias in the statistical properties of the population. Two kinds of errors may appear during generation. *Errors in statistical distribution* may appear because the generated population $\mathcal{A}$ is not large enough given the combinations of attributes' values described by the agent BN. These errors are measured by learning the agent BN on data, and quantified as the average difference between theoretical and measured probability. As shown in Fig. 5 *(left)*, these errors (bottom curve) remain low and are negatively correlated to the population size. *Errors in matching* may appear when no candidate is found to link several agents, and are quantified as the rate of the total number of links required by RC_t on the number of created ties. If this error decreases while the population size increases, then errors are only due to the discrete nature of agents: there will always be a number of agents, which can not be connected because their theoretical peer was not created. As shown in Fig. 5, this error rate quickly drops above a given population size. Given our parameters, a population of 10,000 agents is the minimum to reduce errors. If the matching error rate remains high when the population size increases, it means that agent BN and/or matching BN are incompatible. In Fig. 5, the curve for the *married* link shows that the number of wives per man is not compatible with the proportion of married wives. In this case, statistics (or assumptions) used to build BN should be checked and corrected.

Figure 5 *(right)* depicts the evolution of statistical properties of the relationship network. Density is low (under 0.01). Transitivity (sometimes called clustering rate) is high, and becomes stable above the 10,000 agents threshold. Links defined across spatial locations (for family, work, and with low probability friendship) play the role of shortcuts, so the average path length in the model grows very slowly (around 4.8), exhibiting the so-called "small-world" property. The average degree is in theory defined by attributes RC_t. In fact, it is only reached when all the required links are created (above 10,000 agents), then sticks to its theoretical value.
At evidence, there is a minimum population size to satisfy constraints defined by the matching BN. The more the matching BN are constraining, the higher the threshold. Above this threshold, statistical properties remain remarkably stable.

5 Discussion

In this paper, we proposed a methodology to formalize various statistics available for the population (R5) to generate a simplified network of relationships at a country scale (R1). The resulting network includes agents' attributes (R3) and different kinds of relationships (R2), so the modeler can define with more precision if interaction takes place. The network exhibits a high clustering rate, low density and a low average path length. Formalism enables modelers to comply with evidence on social selection processes (R4) like affiliation, homophily and transitivity. We illustrated the methodology by generating a network of relationships in rural Kenya in which socio-demographic studies, sociological findings, and qualitative observations on affiliation are put together to reproduce familial, work and friendship relationships.
The choice of Bayesian networks to formalize scattered statistics makes the fusion of different statistical sources more intuitive, threfore any social scientist can use this generator. BN also facilitate the generation of a heterogeneous population of agents and the creation of links between these agents.

The purpose of this methodology is to generate a *model* of relationships in a population. Thus, the generated graph is only a *simplification of real relationships given available data*, and do not target the same precision as models at a smaller scale. However, the network generated is rooted in reality by using the

available statistics and observations of the modeled population. In some way, we hope it fills the gap between models proposed by social scientists (highly descriptive, but limited to small populations) and generators physicists (generate large populations with a low descriptive power). We plan to soon publish the software which implements the generator.
We decided to illustrate this paper with social interactions in rural Kenya because of the relative simplicity of its social structure. The next step is to model more complex populations of industrialized countries (more affiliations, socioeconomic classes, attributes). Our agenda of research also includes the *investigation of dynamics* supported by generated networks, especially in the frame of information diffusion, and the *formal analysis of generated networks' properties*.

References

1. Wasserman, S., Faust, K.: Social network analysis, methods and applications. Cambridge: Cambridge University Press (1994)
2. Thiriot, S., Kant, J.D.: Using associative networks to represent adopters' beliefs in a multiagent model of innovation diffusion. Advances in Complex Systems **11**(2) (2008) 261–272
3. Thiriot, S., Kant, J.D.: Reproducing stylized facts of word-of-mouth with a naturalistic multi-agent model. In: Second World Congress on Social Simulation. (2008) to appear.
4. Rogers, E.M.: Diffusion of Innovations. 5th edn. New York: Free Press (2003)
5. Granovetter, M.: The Strength of Weak Ties. The American Journal of Sociology **78**(6) (1973) 1360–1380
6. Robins, G., Elliott, P., Pattison, P.: Network models for social selection processes. Social Networks **23** (2001) 1–30
7. McPherson, M., Smith-Lovin, L., Cook, J.: Birds of a feather: Homophily in social networks. Annual Reviews in Sociology **27** (2001) 415–444
8. KDHS: Kenya demographic and health survey 2003. Central Bureau of Statistics (2003)
9. Mburugu, E.K., Adams, B.N.: Families in Kenya. In: Handbook of World Families. SAGE (2004) 3–24
10. Watkins, S., Rutenberg, N., Green, S.: Diffusion and debate: Controversy about reproductive change in Nyanza Province, Kenya. Annual Meeting of the Population Association of America (1995) 6–8
11. Rutenberg, N., Watkins, S.: The buzz outside the clinics: conversations and contraception in Nyanza Province, Kenya. Studies in Family Planning **28**(4) (1997) 290–307
12. Watkins, S.C., Rutenberg, N., Green, S., Onoko, C., White, K., Franklin, N., Clark, S.: 'circle no bicycle' or everything you should want to know about survey data but were afraid to ask. SNP Working Paper No.1, Philadelphia: University of Pennsylvania (1995)
13. Phan, D., Amblard, F., eds.: Agent-based Modelling and Simulation in the Social and Human Sciences. Oxford, The Bardwell Press (2007)
14. Frank, O., Strauss, D.: Markov graphs. Journal of the American Statistical Association **81** (1986) 832–842
15. Jensen, F.: Introduction to Bayesian Networks. Springer-Verlag New York, Inc. Secaucus, NJ, USA (1996)