Peer Effects in Random Consideration Sets*

Nail Kashaev
Western University
nkashaev@uwo.ca

Natalia Lazzati
UC Santa Cruz
nlazzati@ucsc.edu

This version: September 1, 2020
First version: April 13, 2019, arXiv:1904.06742

Abstract This paper develops a dynamic model of discrete choice that incorporates peer effects into random consideration sets. We characterize the equilibrium behavior and study the empirical content of the model. In our setup, changes in the choices of friends induce changes in the distribution of the consideration sets. We exploit this variation to recover the ranking of preferences, the attention mechanisms, and the network connections. These identification results allow unrestricted heterogeneity across people and do not rely on variation of either covariates or the set of available options. Our methodology leads to a maximum-likelihood estimator that performs well in simulations.

JEL codes: C31, C33, D83, O33
Keywords: Peer Effects, Random Consideration Sets, Continuous Time Markov Process

*For their useful comments and suggestions we would like to thank Victor Aguiar, Roy Allen, Tim Conley, Paola Manzini, Marco Mariotti, and Salvador Navarro. We would also like to thank the audience at Workshop in Experiments, Revealed Preferences, and Decisions, the Inter University Colloquium on Analytical Research in Economics, and the CARE-Colloquium series at University of Kansas.

1. Introduction

In the last few years, the basic rational choice model of decision-making has been modified in many dimensions. This was partly due to the recognition that cognitive factors might play an important role in determining the choices of people. These works have produced, among others, the so-called consideration set models. These models assume that people do not consider all the available options at the moment of choosing but a subset of them. It is still an open question in the literature how the consideration sets are formed and what the main factors in their formation process are. We believe that peer effects could have a key role in determining the subset of options that people pay attention to in many environments. This paper builds from this belief.

We construct a dynamic multinomial choice model where the choices of friends affect the subset of options that each person ends up considering. Due to limited consideration, people might ignore their most preferred alternatives quite often. In our model, the structure of the network of friends affects the strength of these mistakes. We show that all parts of the model can be uniquely identified and estimated from a long sequence of choices. These parts include the ranking of preferences, the attention mechanism (or consideration probabilities), and the set of connections of each person. There are two aspects of our identification results that deserve special attention. First, we allow unrestricted heterogeneity across people with respect to all the parts of the model. Second, in contrast to most other works on consideration sets, we do not rely on variation of either covariates or the set of available options (or menu) to recover these components.

More specifically, in our model, people are linked through a social network. At a randomly given time, a random person gets the opportunity to select a new option out of a finite set of alternatives. The person sticks to her new option until the revision opportunity arises again. As we mentioned earlier, people do not consider all the available options at the moment of revising their selection. Instead, each person first forms a consideration set and then picks her most preferred option from it. A distinctive feature of our model is that the probability that a given alternative enters the consideration set depends on the number of friends that are currently adopting that option. This model leads to a sequence of joint choices that evolve through time according to a continuous-time Markov process.

We initially show the dynamic system has a unique equilibrium. The equilibrium consists of an invariant distribution in the space of joint choices across people in the network. It specifies the fraction of time each profile of joint choices prevails in the long run. We show existence and uniqueness of the equilibrium distribution under the assumption that each option has (a priori) nonzero probability of being considered by each person irrespective of the choices of friends. This assumption captures the idea that a person can eventually pay attention to an alternative for various reasons that are outside the control of our model (e.g., watching an ad on television or receiving a coupon). It assures that we can move from any initial configuration of choices to any other one in finite time. We then show that the model primitives (i.e., the preferences, the attention mechanism, and the network structure) are uniquely identified.

We assume the researcher has access to a long sequence of joint choices. In our dynamic model, these observed choices are generated by a system of conditional choice probabilities; each of these conditional choice probabilities specifies the distribution of choices of a given person conditional on the choices of others (at the moment of revising her selection). The identification strategy we offer is a two-step procedure.

First, we assume the conditional choice probabilities have been recovered from the data. We also assume a person is more likely to pay attention to a specific option if more of her friends are currently adopting it. Thus, changes in the choices of friends induce stochastic variation in consideration sets. We exploit this variation to recover the set of connections between the people in the network and their ranking of preferences. We then use this information to recover the attention mechanism of each person, i.e., the probability of including a specific option in the consideration set as a function of the number of friends who are currently choosing it.

Second, we study identification of the conditional choice probabilities. We consider two datasets: continuous-time data and discrete-time data with arbitrary time intervals. These two datasets coincide in that they provide a long sequence of choices from people in the network. They differ in the timing at which the researcher observes these choices. In continuous-time datasets, the researcher observes people's choices in real time. This allows the researcher to record the precise moment at which a person revises her strategy and the configuration of people's choices at that time. We can think of this dataset as the "ideal dataset." With the proliferation of online platforms and scanner, this sort of data might be available in some applications. In this dataset, the researcher directly recovers the conditional choice probabilities. In discrete-time datasets, the researcher observes the joint configuration of choices at fixed time intervals (e.g., the choice configuration is observed every Monday). This second dataset is less informative than the first one, but we believe it might be more common in practice. In this case, the conditional choice probabilities are not directly observed or recovered, they need to be inferred from the data. Interestingly, adding only a mild extra condition, we show that here the conditional choices are also uniquely identified. For this last result we invoke insights from Blevins (2017, 2018). Briefly, the reason for which the second dataset suffices to recover the primitives of the model lies in the fact that the transition rate matrix of a continuous-time process with independent revision times across people is rather parsimonious. In particular, the probability that two or more people revise their selected options at the same time is zero. This property translates into a transition rate matrix that has zeros in many known locations. The non-zero elements can be then nicely recovered and they constitute a one-to-one mapping with the conditional choice probabilities.

After presenting the main results, we show how our ideas work in practice with a model of restaurant choice. The application clarifies how the structure of the network affects the strength of the mistakes that people make. It shows, for example, that homophily reduces the frequency by which people choose dominated options —here, dominated is with respect to the preference order of each person over the set of all the available options. It also shows that having more friends helps in the same regard. We also use the application to highlight some estimation aspects of our model. In particular, we show that our maximum-likelihood estimator performs well even for relatively small data sets.

Our initial model relies on three main assumptions: (i) the preferences of each person are deterministic, (ii) there exists one alternative (the default) that is picked if and only if nothing else is considered, and (iii) the distribution of consideration sets is multiplicatively separable across alternatives and the probability of including each option depends on the number (but not the identity) of the friends that selected that option. We extend our model in various directions to relax these assumptions. First, we consider random preferences (on top of random consideration sets). The network structure and the attention mechanism are identified without any new assumptions. The random preferences are identified if and only if each person has "enough" friends —the actual number of friends needed depends on the number of options. Second, we analyze a model with no default. All the primitives of this model are identified if there are more than three options in the set of available alternatives. Finally, we consider a set of extensions where consideration sets are formed arbitrarily (e.g., no multiplicative separability assumption). In these extensions, different friends may have different effects on the attention mechanism of a given person. In this model, we still can uniquely recover preferences and the network structure. The consideration probabilities, in general, are partially identified. However, we show that different forms of symmetry between consideration set probabilities are sufficient for their identification.

We finally relate our results with the existing literature. From a modeling perspective, our setup combines the dynamic model of social interactions of Blume (1993, 1995) with the (single-agent) model of random consideration sets of Manzini and Mariotti (2014). By adding peer effects in the consideration sets we can use variation in the choices of others as the main tool to recover preferences. The literature on identification of single-agent consideration set models has mainly relied on variation of the set of available options or menus. The latter includes Aguiar (2017), Aguiar et al. (2016), Brady and Rehbeck (2016), Caplin et al. (2018), Cattaneo et al. (2017), Horan (2019), Lleras et al. (2017), Manzini and Mariotti (2014), Masatlioglu et al. (2012), and Dardanoni et al. (2020).[1] (See Aguiar et al., 2019 for a comparison of several consideration set models in an experimental setting.) Other papers have relied on the existence of exogenous covariates that shift preferences or consideration sets. The latter include Aguiar and Kashaev (2020), Barseghyan et al. (2019), Conlon and Mortimer (2013), Draganska and Klapper (2011), Gaynor et al. (2016), Goeree (2008), Mehta et al. (2003), and Roberts and Lattin (1991). Variation of exogenous covariates has also been used by Abaluck and Adams (2017) via an approach that exploits symmetry breaks with respect to the full consideration set model.

As we mentioned earlier, we can recover from the data the set of connections between the people in the network. In the context of linear models, a few recent papers have made progress in the same direction. Among them, Blume et al. (2015), Bonaldi et al. (2015), De Paula et al. (2018), and Manresa (2013). In the context of discrete-choice, Chambers et al. (2019) also identifies the network structure but in their model peers do not affect consideration sets but directly change preferences (among other differences).

In the paper, we connect the equilibrium behavior of our model with the Gibbs equilibrium. This connection is similar to the one in Blume and Durlauf (2003). We also distinguish our approach with the more standard models of peer effects in preferences. To this end we embed the models of

[1]See also Manski (1977) for a throughout formulation of the discrete choice model that incorporates the possibility that the decision maker only considers a subset of options.

Brock and Durlauf (2001, 2002) into our dynamic revision process.

Let us finally mention two other papers that incorporate peer effects in the formation of consideration sets. Borah and Kops (2018) do so in a static framework and rely on variation of menus for identification. Lazzati (2020) considers a dynamic model but the time is discrete and she focuses on two binary options that can be acquired together.

The rest of the paper is organized as follows. Section 2 presents the model and differentiates it from the more standard model of peer effects in preferences. Section 3 describes the equilibrium behavior. Section 4 studies the empirical content of the model. Section 5 extends the initial idea to contemplate random preferences (besides random consideration sets), the case of no-default option, and more general formation process for the consideration sets. Section 6 presents some simulation and estimation results for a model of choosing a restaurant. Section 7 concludes, and all the proofs are collected in Appendix A. Appendices B and C cover the Gibbs random field model and provide some simulation results, respectively.

2. The Model

This section describes the model and the main assumptions we invoke in the paper. It also relates our approach to the more standard model of peer effects in preferences.

2.1. Social Network, Consideration Sets, and Choices

Network and Choice Configuration There is a finite set of people connected through a social network. The network is described by a simple graph $\Gamma = (\mathcal{A}, e)$, where $\mathcal{A} = \{1, 2, \ldots, A\}$ is the finite set of nodes (or people) and e is the set of edges. Each edge identifies two connected people and the direction of the connection. For each Person $a \in \mathcal{A}$, her set of friends (or reference group) is defined as follows

$$\mathcal{N}_a = \{a' \in \mathcal{A} : a' \neq a \text{ and there is an edge from } a \text{ to } a' \text{ in } e\}.$$

There is a set of alternatives $\overline{\mathcal{Y}} = \mathcal{Y} \cup \{o\}$, where $\mathcal{Y} = \{1, 2, \ldots, Y\}$ is a finite set of options and o is a default option. Each Person a has a strict preference order $\succ_a$ over the set of options $\mathcal{Y}$. The default option is chosen only if nothing else is considered.[2] (We extend the analysis to random preferences in Section 5.1 and relax the specification of the default option in Section 5.2.) We refer to $\mathbf{y} = (y_a)_{a \in \mathcal{A}} \in \overline{\mathcal{Y}}^A$ as a choice configuration.

[2]Equivalently we can assume that the default option is always considered and it is the worst alternative for every person. Moreover, though we assume the same default option for all people, all our results would go through if people have different default options as far as the researcher knew the identity of the worst option for each person.

Choice Revision We model the revision of choices as a standard continuous-time Markov process. In particular, we assume that people are endowed with independent Poisson alarm clocks with rates $\lambda = (\lambda_a)_{a\in\mathcal{A}}$.[3] At randomly given moments (exponentially distributed with mean $1/\lambda_a$) the alarm of Person a goes off.[4] When this happens, the person selects the most preferred alternative among the ones she is actually considering. Formally, if $\mathcal{C} \subseteq \mathcal{Y}$ is her consideration set, then the choice of Person a can be represented by an indicator function

$$\mathrm{R}_a(v \mid \mathcal{C}) = \mathbb{1}\left(v \succ_a v' \text{ for all } v' \in \mathcal{C} \text{ and } v \in \mathcal{C}\right)$$

that takes value 1 if v is the most preferred option in $\mathcal{C}$ according to $\succ_a$. If, at the moment of choosing, the consideration set of Person a does not include any alternative in $\mathcal{Y}$, then the person simply selects the default option.

Peer Effects in the Formation of Consideration Sets In our model, whether Person a pays attention to a particular alternative depends on her own choice and the configuration of choices of her friends at the moment of revising her selection. We indicate by $\mathrm{Q}_a(v \mid \mathbf{y})$ the probability that Person a pays attention to alternative v given a choice configuration $\mathbf{y}$. It follows that the probability of facing consideration set $\mathcal{C}$ has the form of

$$\prod\nolimits_{v\in\mathcal{C}} \mathrm{Q}_a(v \mid \mathbf{y}) \prod\nolimits_{v\notin\mathcal{C}} \left(1 - \mathrm{Q}_a(v \mid \mathbf{y})\right).$$

By combining preferences and stochastic consideration sets, the probability that Person a selects (at the moment of choosing) alternative $v \in \mathcal{Y}$ is given by

$$\mathrm{P}_a(v \mid \mathbf{y}) = \mathrm{Q}_a(v \mid \mathbf{y}) \prod\nolimits_{v'\in\mathcal{Y}, v'\succ_a v} \left(1 - \mathrm{Q}_a(v' \mid \mathbf{y})\right). \tag{1}$$

The probability of selecting the default option o is just $\prod_{v\in\mathcal{Y}} \left(1 - \mathrm{Q}_a(v \mid \mathbf{y})\right)$. Leaving aside peer effects, this process of formation of consideration sets corresponds to the one analysed in Manzini and Mariotti (2014). In Section 5.3 we extend our analysis to a more general setting.

Altogether, the above three components characterize our dynamic model of peer effects in random consideration sets. This model leads to a sequence of choices that evolves through time according to a Markov random process.

Let us add a few final comments about our model. First, it represents truly boundedly rational agents. The people in our framework do not solve a dynamic optimization problem and their choice sets may thereby not include their most preferred alternatives for long periods of time. Second, in our initial specification, the only source of randomness in choice is via consideration sets. In this sense, our initial model captures a single, though important, channel of possible mistakes in choices. The social network shapes the nature and the strength of these mistakes. As we mentioned

[3] See Blume (1993, 1995) for theoretical models that rely on Poisson alarm clocks and Blevins (2018) for a nice discussion of the advantages of this type of revision process from an applied perspective.

[4] That is, each Person a is endowed with a collection of random variables $\{\tau_n^a\}_{n=1}^{\infty}$ such that each difference $\tau_n^a - \tau_{n-1}^a$ is exponentially distributed with mean $1/\lambda_a$. These differences are independent across people and time.

earlier, we extend the analysis to random preferences in Section 5.1. In this extension, $\mathrm{R}_a(\cdot \mid \mathcal{C})$ is not an indicator function but a distribution on $\mathcal{Y}$. Lastly, notice that we modeled peer effects via consideration sets. In this sense, our approach departs from the canonical model of peer effects in preferences. We add more to this distinction below.

2.2. Main Assumptions

Our results build on three assumptions. Let $\mathrm{N}_a^v(\mathbf{y})$ be the number of friends of Person a who select option v in choice configuration $\mathbf{y}$. Formally,

$$\mathrm{N}_a^v(\mathbf{y}) = \sum\nolimits_{a' \in \mathcal{N}_a} \mathbb{1}(y_{a'} = v).$$

We indicate by $|\mathcal{N}_a|$ the cardinality of $\mathcal{N}_a$. The three assumptions are as follows.

(A1) For each $a \in \mathcal{A}$, $v \in \mathcal{Y}$, and $\mathbf{y} \in \overline{\mathcal{Y}}^A$, $1 > \mathrm{Q}_a(v \mid \mathbf{y}) > 0$.

(A2) For each $a \in \mathcal{A}$, $|\mathcal{N}_a| > 0$.

(A3) For each $a \in \mathcal{A}$, $v \in \mathcal{Y}$, and $\mathbf{y} \in \overline{\mathcal{Y}}^A$,

$$\mathrm{Q}_a(v \mid \mathbf{y}) \equiv \mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})) \text{ is strictly increasing in } \mathrm{N}_a^v(\mathbf{y}).$$

Assumption A1 states that, for any choice configuration, the probability of considering each option is strictly positive and lower than one, independently on how many friends have selected that option. This assumption captures the idea that a person can eventually pay attention to an alternative for various reasons that are outside the control of our model (e.g., watching an ad on television or receiving a coupon). It also allows the person to eventually disregard any further consideration of a given option, including the one that she is currently adopting. From a technical perspective, it follows from A1 that each subset of options is (ex-ante) considered with nonzero probability. Assumption A2 requires each person to have at least one friend. Assumption A3 states that the probability that a given person pays attention to a specific option depends on the current choice of the person and the number (but not the identity) of friends that currently selected it. We relax this specification in Section 5.3. Assumption A3 also states that each person pays more attention to a particular option if more of her friends are adopting it.

2.3. Relation with Peer Effects in Random Preferences

In the same dynamic revision process, we could embed a standard model of peer effects in preferences (instead of in consideration sets) by modifying the conditional choice probabilities. Following, for instance, the canonical models of Brock and Durlauf (2001, 2002) we can assume that the payoff Person a gets from alternative v is the sum of a deterministic private utility the person receives from the choice, $\delta_{a,v}$, a social component that depends on the number of friends that select that option,

$\mathrm{S}_{a,v}(\mathbf{y}) \equiv \mathrm{S}_{a,v}(y_a, \mathrm{N}_a^v(\mathbf{y}))$, and an idiosyncratic term that is doubly exponentially distributed with index parameter 1. Let us also assume that people consider all available alternatives at the moment of revising strategies and that the utility from the default option is normalized to 0. In this case, the probability that an option v is chosen given network configuration $\mathbf{y}$ is given by

$$\mathrm{P}_a(v \mid \mathbf{y}) = \frac{\exp\left(\delta_{a,v} + \mathrm{S}_{a,v}(y_a, \mathrm{N}_a^v(\mathbf{y}))\right)}{1 + \sum_{v' \in \mathcal{Y}} \exp\left(\delta_{a,v'} + \mathrm{S}_{a,v'}\left(y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right)\right)}.$$

Recall that in our consideration set model, the analogous expression takes the form of

$$\mathrm{P}_a(v \mid \mathbf{y}) = \mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})) \prod_{v' \in \mathcal{Y}, v' \succ_a v} \left(1 - \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right)\right).$$

These two models have different empirical implications. In particular, in the random consideration set model, the probability that Person a selects alternative v does not vary if some of her friends switch from the default option to any option that is dominated by alternative v according to $\succ_a$. In contrast, in the model of peer effects in preferences, the probability of selecting alternative v will strictly increase under the same scenario. It follows that (if we can recover conditional choice probabilities from the data, as we do in Section 4) the two models can be set apart.

From a deeper economical perspective, the two models also differ regarding the role played by peers: while both models induce similar behavior among friends with (a priori) different preferences, they do so in a rather different way. The model of peer effects in preferences induces similar choices by decreasing the differences in payoffs among friends. Thus, homophily in preferences arises endogenously in this model. The model of peer effects in consideration sets induces similar choices by making friends pay attention to similar subsets of options. In this model, differences in preferences remain unchanged; the structure of peers simply affects the probability of making mistakes (i.e., the possibility of not selecting the best option because it is just not considered). Thus, homophily in preferences is exogenous to the consideration set model and it shapes the nature and the strength of the mistakes that people make. This distinction matters for the design of policies that aim, for instance, to increase the revenues from selling a specific good. While it may be hard to influence peoples' preferences, it may be easy to increase peoples' awareness of certain alternatives.

3. Equilibrium Behavior

3.1. Equilibrium

The independent identically distributed Poisson alarm clocks, which lead the selection revision process, guarantee that at each moment of time at most one person revises her selection almost surely. Thus, the transition rates between choice configurations that differ in more than one person changing her selection are zero. The advantage of this fact for model identification is rather clear: there are fewer terms to recover. Blevins (2017, 2018) discuss this feature and its advantages over discrete time models. Formally, the transition rate from choice configuration $\mathbf{y}$ to any different one $\mathbf{y}'$ is as follows

$$\mathrm{m}\left(\mathbf{y}' \mid \mathbf{y}\right) = \begin{cases} 0 & \text{if } \sum_{a\in\mathcal{A}} \mathbb{1}\left(y'_a \neq y_a\right) > 1 \\ \sum_{a\in\mathcal{A}} \lambda_a \, \mathrm{P}_a\left(y'_a \mid \mathbf{y}\right) \mathbb{1}\left(y'_a \neq y_a\right) & \text{if } \sum_{a\in\mathcal{A}} \mathbb{1}\left(y'_a \neq y_a\right) = 1 \end{cases}. \tag{2}$$

In the statistical literature on continuous-time Markov processes these transition rates are the out of diagonal terms of the *transition rate matrix* (also known as the *infinitesimal generator matrix*). The diagonal terms are simply build from these other values as follows

$$\mathrm{m}\left(\mathbf{y} \mid \mathbf{y}\right) = -\sum\nolimits_{\mathbf{y}'\in\overline{\mathcal{Y}}^A\setminus\{\mathbf{y}\}} \mathrm{m}\left(\mathbf{y}' \mid \mathbf{y}\right).$$

We will indicate by $\mathcal{M}$ the transition rate matrix. In our model, the number of choice configurations is $(Y+1)^A$. Thus, $\mathcal{M}$ is a $(Y+1)^A \times (Y+1)^A$ matrix. There are many different ways of ordering the choice configurations and thereby writing the transition rate matrix. To avoid any sort of ambiguity in the exposition, we will let the choice configurations be ordered according to the lexicographic order with o treated as zero. Constructed in this way the first element of $\mathcal{M}$ is $\mathcal{M}_{11} = \mathrm{m}\left((o,o,\ldots,o)' \mid (o,o,\ldots,o)'\right)$. Formally, let $\iota\left(\mathbf{y}\right) \in \left\{1,2,\ldots,(Y+1)^A\right\}$ be the position of $\mathbf{y}$ according to the lexicographic order. Then,

$$\mathcal{M}_{\iota(\mathbf{y})\iota(\mathbf{y}')} = \mathrm{m}\left(\mathbf{y}' \mid \mathbf{y}\right).$$

An equilibrium in our model is an invariant distribution $\mu : \overline{\mathcal{Y}}^A \to [0,1]$, with $\sum_{\mathbf{y}\in\overline{\mathcal{Y}}^A} \mu\left(\mathbf{y}\right) = 1$, of the dynamic process with transition rate matrix $\mathcal{M}$. It indicates the likelihood of each choice configuration $\mathbf{y}$ in the long run. This equilibrium behavior relates to the transition rate matrix in a linear fashion

$$\mu\mathcal{M} = \mathbf{0}.$$

Proposition 3.1 states equilibrium existence and uniqueness for our model.

Proposition 3.1. *If Assumption A1 is satisfied, then there exists a unique* μ.

The next example describes the equilibrium behavior of a simple specification of our model.

Example 1: There are two identical, connected people that select between two alternatives, namely, option 1 and the default option o. The rates for their Poisson alarm clocks are 1. We will also assume, to simplify the setup, that the probability of paying attention to a particular option only depends on the current choice of the other person. Thus, for $a = 1, 2$, we get that

$$\mathrm{P}_a\left(1 \mid \mathbf{y}\right) = \mathrm{Q}\left(1 \mid \mathrm{N}^v_a\left(\mathbf{y}\right)\right) \text{ and } \mathrm{P}_a\left(o \mid \mathbf{y}\right) = 1 - \mathrm{Q}\left(1 \mid \mathrm{N}^v_a\left(\mathbf{y}\right)\right).$$

Note that we avoided the subindex in Q because of the symmetry and dropped the dependence on the previous choice because of our simplifying assumption.

The transition rate matrix $\mathcal{M}$ is as follows. (The columns are ordered as the rows.)

(o,o)	$-2\,\mathrm{Q}(1 \mid 0)$	$\mathrm{Q}(1 \mid 0)$	$\mathrm{Q}(1 \mid 0)$	0
$(o,1)$	$1-\mathrm{Q}(1 \mid 0)$	$-1+\mathrm{Q}(1 \mid 0)-\mathrm{Q}(1 \mid 1)$	0	$\mathrm{Q}(1 \mid 1)$
$(1,o)$	$1-\mathrm{Q}(1 \mid 0)$	0	$-1+\mathrm{Q}(1 \mid 0)-\mathrm{Q}(1 \mid 1)$	$\mathrm{Q}(1 \mid 1)$
$(1,1)$	0	$1-\mathrm{Q}(1 \mid 1)$	$1-\mathrm{Q}(1 \mid 1)$	$-2+2\,\mathrm{Q}(1 \mid 1)$

.

The transition rate matrix $\mathcal{M}$ is naturally more complex when there are more actions or more people. However, the structure of the zeros in $\mathcal{M}$ is rather similar. As we mentioned earlier, this feature of the model facilitates identification and estimation.

The invariant distribution of choices, or steady-state equilibrium, satisfies

$$\mu\mathcal{M} = \mathbf{0}.$$

Solving this system of equations, we get that the steady-state equilibrium is

$$\begin{aligned}
\mu(o,o) &= \frac{[1-\mathrm{Q}(1 \mid 0)]\,[1-\mathrm{Q}(1 \mid 1)]}{1-\mathrm{Q}(1 \mid 1)+\mathrm{Q}(1 \mid 0)} \\
\mu(o,1) = \mu(1,o) &= \frac{\mathrm{Q}(1 \mid 0)\,[1-\mathrm{Q}(1 \mid 1)]}{1-\mathrm{Q}(1 \mid 1)+\mathrm{Q}(1 \mid 0)} \\
\mu(1,1) &= \frac{\mathrm{Q}(1 \mid 0)\,\mathrm{Q}(1 \mid 1)}{1-\mathrm{Q}(1 \mid 1)+\mathrm{Q}(1 \mid 0)}.
\end{aligned}$$

The steady-state equilibrium is a joint distribution on the pair of choice configurations. It states the fraction of time that each pair of choices (y_1, y_2) prevails in the long run. ■

3.2. Relation to Gibbs Equilibrium

We finish this section by connecting our setup with the so-called Gibbs random field models. This is an interesting exercise as these models have been widely used in Economics to study social interactions. (See Allen, 1982, Blume, 1993, 1995, and Blume and Durlauf, 2003, among many others.)

The starting point of the Gibbs random field models is a set of conditional probability distributions. Each element of the set describes the probability that a given person selects each alternative as a function of the profile of choices of the other people. In our model, the set of conditional probabilities is $(\mathrm{P}_a)_{a\in\mathcal{A}}$, with a generic element given by

$$\mathrm{P}_a(v \mid \mathbf{y}) = \mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})\right) \prod\nolimits_{v'\in\mathcal{Y}, v'\succ_a v}\left(1-\mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right)\right) \text{ for } v \in \mathcal{Y} \text{ and } \mathbf{y} \in \overline{\mathcal{Y}}^A.$$

A Gibbs equilibrium is defined as a joint distribution over the vector of choices $\mathbf{y}$, $\mathrm{P}(\mathbf{y})$, that is able to generate $(\mathrm{P}_a)_{a\in\mathcal{A}}$ as its conditional distribution functions.

Gibbs equilibria typically do not exist. (In the statistical literature, a similar existence problem is referred as the issue of compatibility of conditional distributions.) The existence of Gibbs equilibria depends on a great deal of homogeneity among people. When such a joint distribution exists, then it coincides with the invariant distribution μ of the dynamic revision process of our model. We prove this claim in Appendix B and illustrate the idea with Example 1. (See Blume and Durlauf (2003) for a nice discuss.)

Example 1 (continued): The conditional distributions of choices in Example 1 satisfy the compatibility requirements. Thus, there exists a joint distribution on y_1 and y_2 that can generate the conditional distributions of choices as its conditional distribution functions. In this simple case, the invariant distribution μ of the dynamic revision process coincides with the Gibbs equilibrium of the model. To see this notice that from our previous results we get

$$\begin{aligned}\mu(o) &= \mu(o,o)+\mu(o,1)=\mu(o,o)+\mu(1,o)=\frac{1-\mathrm{Q}(1\mid 1)}{1-\mathrm{Q}(1\mid 1)+\mathrm{Q}(1\mid 0)},\\ \mu(1) &= \mu(1,1)+\mu(o,1)=\mu(1,1)+\mu(1,o)=\frac{\mathrm{Q}(1\mid 0)}{1-\mathrm{Q}(1\mid 1)+\mathrm{Q}(1\mid 0)}.\end{aligned}$$

(Here again, we avoided subindices due to the symmetry across people.) Thus, the conditional distributions are

$$\begin{aligned}\mu(1,o)/\mu(o) &= \mathrm{Q}(1\mid 0)\text{ and }\mu(o,o)/\mu(o)=1-\mathrm{Q}(1\mid 0)\\ \mu(1,1)/\mu(1) &= \mathrm{Q}(1\mid 1)\text{ and }\mu(o,1)/\mu(1)=1-\mathrm{Q}(1\mid 1).\end{aligned}$$

It follows that the pair of conditional probabilities generated by equilibrium behavior coincide with the conditional distributions of choices in the model. ■

4. Empirical Content of the Model

This section provides conditions under which the researcher can uniquely recover (from a long sequence of choices) the set of connections $\Gamma=(\mathcal{A},e)$, the profile of strict preferences $\succ=(\succ_a)_{a\in\mathcal{A}}$, the attention mechanism $\mathrm{Q}=(\mathrm{Q}_a)_{a\in\mathcal{A}}$, and the rates of the Poisson alarm clocks $\lambda=(\lambda_a)_{a\in\mathcal{A}}$. We offer alternative conditions under which the model is identified. In this section, as it is the case with identification, we will abstract from small sample issues.

We separate the identification analysis in two parts. Let $\mathrm{P}=(\mathrm{P}_a)_{a\in\mathcal{A}}$ be the profile of choice probabilities of people in the network. Each $\mathrm{P}_a(v\mid\mathbf{y}):\overline{\mathcal{Y}}\times\overline{\mathcal{Y}}^A\to(0,1)$ specifies the (ex-ante) probability that Person a selects option v when the choice configuration is $\mathbf{y}$. Recall that, in our model,

$$\mathrm{P}_a(v\mid\mathbf{y})=\mathrm{Q}_a(v\mid\mathbf{y})\prod\nolimits_{v'\in\mathcal{Y},v'\succ_a v}(1-\mathrm{Q}_a(v'\mid\mathbf{y})),$$

and the probability of selecting the default option o is $\prod_{v\in\mathcal{Y}}\left(1-\mathrm{Q}_a\left(v\mid\mathbf{y}\right)\right)$. First, we show that each set of conditional choice probabilities P maps into a unique set of connections, profile of strict preferences, and attention mechanism. Thus, knowledge of the set of the conditional choice probabilities allows us to uniquely recover all the elements of the model. Second, we build identification of the conditional choice probabilities P from a long sequence of choices.

4.1. Identification of the Model from P

Under Assumptions A1-A3, changes in the choices of friends induce stochastic variation of the consideration sets. Also, A3 guarantees this variation is monotonic in the sense that the probability of considering one option increases with the number of friends that are currently adopting it. This variation allows us to recover the set of connections between the people in the network and the ranking of preferences of each of them. We then sequentially identify the attention mechanism of each person moving from the most preferred alternative to the least preferred one. Proposition 4.1 formalizes these claims.

Proposition 4.1. *Under Assumptions A1-A3, the set of connections* Γ*, the profile of strict preferences* $\succ$*, and the attention mechanism* Q *are point identified from* P*.*

The next example sheds light on the identification strategy in Proposition 4.1.

Example 2: Suppose there are three people $\mathcal{A}=\{1,2,3\}$ that select between two alternatives $\mathcal{Y}=\{1,2\}$ and the default option o. The researcher knows P_1, P_2, and P_3. Let us consider Person 1. Let $\mathbf{y}$ be such that $y_1=o$. The probability that Person 1 selects the default option o (given a profile of choices $\mathbf{y}$ with $y_1=o$) is

$$\mathrm{P}_1\left(o\mid\mathbf{y}\right)=\left(1-\mathrm{Q}_1\left(1\mid o,\mathrm{N}_1^1\left(\mathbf{y}\right)\right)\right)\left(1-\mathrm{Q}_1\left(2\mid o,\mathrm{N}_1^2\left(\mathbf{y}\right)\right)\right).$$

Under A3, we get that $2\in\mathcal{N}_1$ if and only if

$$\mathrm{P}_1\left(o\mid o,o,o\right)>\mathrm{P}_1\left(o\mid o,1,o\right).$$

Indeed, if $2\in\mathcal{N}_1$, then the probability of choosing the default option by Person 1 should decrease if Person 2 picks something else. Also, if $2\notin\mathcal{N}_1$, then the probability of choosing the default option by Person 1 should be invariant to choices of Person 2. Similarly, $3\in\mathcal{N}_1$ if and only if $\mathrm{P}_1\left(o\mid o,o,o\right)>\mathrm{P}_1\left(o\mid o,o,1\right)$. Thus, we can learn (from the data) the set of friends of Person 1. Let us assume that $\mathcal{N}_1=\{2\}$. To recover the preferences of Person 1 note that

$$\begin{array}{ll}\mathrm{P}_1\left(1\mid\mathbf{y}\right)=\mathrm{Q}_1\left(1\mid o,\mathrm{N}_1^1\left(\mathbf{y}\right)\right) & \text{if}\quad 1\succ_1 2\\ \mathrm{P}_1\left(1\mid\mathbf{y}\right)=\mathrm{Q}_1\left(1\mid o,\mathrm{N}_1^1\left(\mathbf{y}\right)\right)\left(1-\mathrm{Q}_1\left(2\mid o,\mathrm{N}_1^2\left(\mathbf{y}\right)\right)\right) & \text{if}\quad 2\succ_1 1\end{array}.$$

Thus, $2\succ_1 1$ if and only if

$$\mathrm{P}_1\left(1\mid o,o,o\right)>\mathrm{P}_1\left(1\mid o,2,o\right).$$

Suppose that, indeed, we get that $2 \succ_1 1$. We can finally recover the attention mechanism (for $y_1 = o$) via the next four probabilities in the data

$$\begin{array}{ll} \mathrm{P}_1(2 \mid o, o, o) = \mathrm{Q}_1(2 \mid o, 0) & \mathrm{P}_1(2 \mid o, 2, o) = \mathrm{Q}_1(2 \mid o, 1) \\ \mathrm{P}_1(1 \mid o, o, o) = \mathrm{Q}_1(1 \mid o, 0)(1 - \mathrm{Q}_1(2 \mid o, 0)) & \mathrm{P}_1(1 \mid o, 1, o) = \mathrm{Q}_1(1 \mid o, 1)(1 - \mathrm{Q}_1(2 \mid o, 0)) \end{array}.$$

By considering two other choice profiles $\mathbf{y}$ with $y_1 = 1$ and $y_1 = 2$ (instead of $y_1 = o$), respectively, we can fully recover the attention mechanism of Person 1. By a similar exercise we can recover the sets of friends, preferences, and the attention mechanisms for Persons 2 and 3. ■

4.2. Identification of P

This section studies identification of the conditional choice probabilities and the rates of the Poisson alarm clocks from two different datasets. These two datasets coincide in that they consider long sequences of choices from people in the network. They differ in the timing at which the researcher observes these choices. In Dataset 1 people's choices are observed in real time. This allows the researcher to record the precise moment at which a person revises her strategy and the configuration of choices at that time. In Dataset 2 the researcher simply observes the joint configuration of choices at fixed time intervals.

Let us assume the researcher observes people's choices at time intervals of length Δ and can consistently estimate $\Pr\left(\mathbf{y}^{t+\Delta} = \mathbf{y}' \mid \mathbf{y}^t = \mathbf{y}\right)$ for each pair $\mathbf{y}', \mathbf{y} \in \overline{\mathcal{Y}}^A$. We will capture these transition probabilities by a matrix $\mathcal{P}(\Delta)$. (Here again, we will assume that the choice configurations are ordered according to the lexicographic order when we construct $\mathcal{P}(\Delta)$.) The connection between $\mathcal{P}(\Delta)$ and the transition rate matrix $\mathcal{M}$ described in Equation (2) is given by

$$\mathcal{P}(\Delta) = e^{(\Delta \mathcal{M})},$$

where $e^{(\Delta \mathcal{M})}$ is the matrix exponential of $\Delta \mathcal{M}$. The two datasets we consider differ regarding Δ: in Dataset 1 we let the time interval be very small. This is an ideal dataset that registers people's choices at the exact time in which any given person revises her choice. As we mentioned earlier, with the proliferation of online platforms and scanner this sort of data might indeed be available for some applications. In Dataset 2 we allow the time interval to be of arbitrary size. The next table formally describes Datasets 1 and 2

Dataset 1	The researcher knows $\lim_{\Delta \to 0} \mathcal{P}(\Delta)$
Dataset 2	The researcher knows $\mathcal{P}(\Delta)$

In both cases, the identification question is whether (or under what extra restrictions) it is possible to uniquely recover $\mathcal{M}$ from $\mathcal{P}(\Delta)$. The first result in this section is as follows.

Proposition 4.2 (Dataset 1)**.** *The conditional choice probabilities* P *and the rates of the Poisson alarm clocks* λ *are identified from Dataset 1.*

The proof of Proposition 4.2 relies on the fact that when the time interval between the observations goes to zero, then we can recover $\mathcal{M}$. There are at least two well-known cases that produce the same outcome without assuming $\Delta \to 0$. One of them requires the length of the interval Δ to be below a threshold $\overline{\Delta}$. The main difficulty of this identification approach is that the value of the threshold depends on the details of the model that are unknown to the researcher. The second case requires the researcher to observe the dynamic system at two different intervals Δ_1 and Δ_2 that are not multiples of each other. (See, e.g., Blevins, 2017 and the literature therein.)

The next proposition states that, by adding an extra restriction, the transition rate matrix can be identified from people's choices even if these choices are observed at the endpoints of discrete time intervals. In this case, the researcher needs to know the rates of the Poisson alarm clocks or normalize them in empirical work.

Proposition 4.3 (Dataset 2)**.** *If Assumption A2 is satisfied, the researcher knows λ, and $\mathcal{M}$ has distinct eigenvalues that do not differ by an integer multiple of $2\pi i/\Delta$, where i here denotes the imaginary unit, then the conditional choice probabilities* P *are generically identified from Dataset 2.*

The key element in proving Proposition 4.3 is that the transition rate matrix of our model is rather parsimonious. To see why, recall that, at any given time, only one person revises her selection with nonzero probability. This feature of the model translates into a transition rate matrix $\mathcal{M}$ that has many zeros in known locations. (See Example 1.) This interesting feature of the model is not shared by models of discrete-time revision processes, where people in the network revise choices simultaneously at fixed time intervals.

5. Extensions of the Model

5.1. Random Preferences

This section extends the initial model to allow randomness in preferences and in consideration sets. In this case, the choice rule $\mathrm{R}_a(\cdot \mid \mathcal{C})$ from Section 2 is not an indicator function but a distribution on $\mathcal{Y}$. We naturally let $\mathrm{R}_a(v \mid \mathcal{C}) = 0$ if $v \notin \mathcal{C}$.

Keeping unchanged the other parts of the model, the probability that Person a selects (at the moment of choosing) alternative $v \in \mathcal{Y}$ is given by

$$\mathrm{P}_a(v \mid \mathbf{y}) = \sum_{\mathcal{C} \subseteq 2^{\mathcal{Y}}} \mathrm{R}_a(v \mid \mathcal{C}) \prod_{v' \in \mathcal{C}} \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right) \prod_{v' \notin \mathcal{C}} \left(1 - \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right)\right). \quad (3)$$

The probability of selecting the default option o is (as before) $\prod_{v \in \mathcal{Y}} (1 - \mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})))$. The next example illustrates the random choice rule with the well-known logit model.

Example 3: If we use the logit model to represent the random preferences of Person a, then the probability that the person selects alternative 1 when alternative 2 is also part of her consideration set would be given by

$$\mathrm{R}_a\left(1 \mid \{1,2\}\right) = \frac{\exp\left(\mathrm{U}_a^1\right)}{\exp\left(\mathrm{U}_a^1\right) + \exp\left(\mathrm{U}_a^2\right)}.$$

In this expression, U_a^1 and U_a^2 are the mean expected utilities that Person a gets from alternatives 1 and 2, respectively. ■

Under this variant of the initial model, the identification of P follows from the same arguments. We will thereby focus on recovering the set of connections, the choice rule, and the attention mechanism from P. The main result is as follows.

Proposition 5.1. *Suppose Assumptions A1-A3 are satisfied. Then, the set of connections* Γ *and the attention mechanism* Q *are point identified from* P. *For each* $a \in \mathcal{A}$, *the random preferences* R_a *are also point identified if, and only if, in addition, we have that* $|\mathcal{N}_a| \geq Y - 1$.

Remark. The last result extends to the case in which the random preferences include the default option o with only one caveat. In this case, the attention mechanism can be recovered up to ratios of the form $\mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v\left(\mathbf{y}\right)\right) / \mathrm{Q}_a\left(v \mid y_a, 0\right)$. That is, we can only recover how much *extra* attention a person pays to each option as more of her friends select that option.

As in our previous results, under Assumptions A1-A3, observed variation in the choices of friends induce stochastic variation of the consideration sets and this variation suffices to recover the connections between the people in the network and the attention mechanism. The only difference with respect to the case of deterministic preferences is that with random preferences we need a larger number of friends for each person, i.e., $|\mathcal{N}_a| \geq Y - 1$ for each $a \in \mathcal{A}$. The extra condition guarantees the matrix of coefficients for the R_a's in expression (3) is full column rank. Indeed, we show that $|\mathcal{N}_a| \geq Y - 1$ is not only sufficient, but necessary, to this end. We illustrate the last result by a simple extension of Example 2 above.

Example 2 (continued): Let us keep all the structure of Example 2 except for people's preferences, which we now assume are random. The identification of the set of connections and the attention mechanism follows from similar ideas. Thus, we will only focus on recovering R_1, R_2, and R_3. Consider the following system of equations for Person 1

$$\begin{pmatrix} \mathrm{P}_1\left(1 \mid y_1, o, o\right) / \mathrm{Q}_1\left(1 \mid y_1, 0\right) \\ \mathrm{P}_1\left(1 \mid y_1, 2, o\right) / \mathrm{Q}_1\left(1 \mid y_1, 0\right) \end{pmatrix} = \begin{pmatrix} 1 - \mathrm{Q}_1\left(2 \mid y_1, 0\right) & \mathrm{Q}_1\left(2 \mid y_1, 0\right) \\ 1 - \mathrm{Q}_1\left(2 \mid y_1, 1\right) & \mathrm{Q}_1\left(2 \mid y_1, 1\right) \end{pmatrix} \begin{pmatrix} \mathrm{R}_1\left(1 \mid \{o, 1\}\right) \\ \mathrm{R}_1\left(1 \mid \{o, 1, 2\}\right) \end{pmatrix}$$

The fact that $\mathrm{R}_1\left(1 \mid \{1\}\right)$ and $\mathrm{R}_1\left(1 \mid \{1,2\}\right)$ can be recovered follows because, by Assumption A3, we have that

$$\det\begin{pmatrix} 1 - \mathrm{Q}_1\left(2 \mid y_1, 0\right) & \mathrm{Q}_1\left(2 \mid y_1, 0\right) \\ 1 - \mathrm{Q}_1\left(2 \mid y_1, 1\right) & \mathrm{Q}_1\left(2 \mid y_1, 1\right) \end{pmatrix} = \mathrm{Q}_1\left(2 \mid y_1, 1\right) - \mathrm{Q}_1\left(2 \mid y_1, 0\right) > 0.$$

The extra condition, $|\mathcal{N}_a| \geq Y - 1$, and Assumption A3 guarantee that the matrix of coefficients for the R_a's is always full column rank. ■

5.2. No Default Option

In the initial model the default option plays a special role: it is chosen if, and only if, nothing else is considered. In some settings, such default option may not exist.[5] This section offers a variant of the model that accommodates to this possibility.[6]

Let us assume there is no default option o, so that $\overline{\mathcal{Y}} = \mathcal{Y}$. The formation process of the consideration set is as before except that, since there is no default option, we need to specify what people do when the consideration set is empty. Given the dynamic nature of our model, we will simply assume that each person sticks to her previous choice if no alternative receives further consideration. Formally, the probability that Person a selects (at the moment of choosing) alternative $v \in \mathcal{Y}$ is given by

$$\mathrm{P}_a(v \mid \mathbf{y}) = \mathrm{Q}_a(v \mid \mathbf{y}) \prod\nolimits_{v' \in \mathcal{Y}, v' \succ_a v} (1 - \mathrm{Q}_a(v' \mid \mathbf{y})) + 1(v = y_a) \prod\nolimits_{v' \in \mathcal{Y}} (1 - \mathrm{Q}_a(v' \mid \mathbf{y})). \tag{4}$$

Proposition 5.2. *Suppose that Assumptions A1-A3 are satisfied and $|\mathcal{Y}| \geq 3$. Then, the set of connections Γ, the profile of strict preferences $\succ$, and the attention mechanism Q are point identified from P.*

5.3. More General Peer Effects in Consideration Sets

In our initial model, the probability that Person a faces consideration set $\mathcal{C}$ given a choice configuration $\mathbf{y}$ takes the form of

$$\prod\nolimits_{v \in \mathcal{C}} \mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})) \prod\nolimits_{v \notin \mathcal{C}} (1 - \mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y}))).$$

This approach entails a multiplicative separable specification of alternatives in the consideration sets. It also assumes that the probability of considering an option depends only on the total number of friends who pick that option, but not on the identity of these friends (i.e., the effect of friends' choices is symmetric across friends). We next show that neither of these assumptions is essential for our approach. Indeed, all the insights previously developed can be used here to extend the initial results.

For each Person a and configuration $\mathbf{y}$, let $\eta_a(\cdot \mid \mathbf{y})$ be an attention index function from $2^{\mathcal{Y}}$ to the positive reals. The value $\eta_a(\mathcal{C} \mid \mathbf{y})$ captures the attention that Person a pays to the set of alternatives $\mathcal{C} \in 2^{\mathcal{Y}}$ given the choice configuration $\mathbf{y}$. The attention-index measures how enticing a consideration set is (see Aguiar et al. (2019) for further details). We define the probability of facing consideration set $\mathcal{C}$ as

$$\frac{\eta_a(\mathcal{C} \mid \mathbf{y})}{\sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} \mid \mathbf{y})}.$$

[5] Also, see Horan (2019).

[6] For alternative ways to close the model see, for example, Barseghyan et al. (2019).

This model corresponds to the consideration set probabilities in Brady and Rehbeck (2016). Since η_a can be identified only up to scale, we normalize the attention-index for the empty set to be 1 (i.e., $\eta(\emptyset \mid \mathbf{y}) = 1$ for all $\mathbf{y} \in \overline{\mathcal{Y}}^A$).[7] This more general setting covers our initial model, which is based on Manzini and Mariotti (2014), as a special case.

By combining preferences and this specification of stochastic consideration sets, the probability that Person a selects (at the moment of choosing) alternative $v \in \mathcal{Y}$ is given by

$$\mathrm{P}_a(v \mid \mathbf{y}) = \sum\nolimits_{\mathcal{C} \in 2^{\mathcal{Y}}: v \in \mathcal{C}} \mathrm{R}_a(v \mid \mathcal{C}) \frac{\eta_a(\mathcal{C} \mid \mathbf{y})}{\sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} \mid \mathbf{y})}.$$

The probability of selecting the default option o is just $\eta_a(\emptyset \mid \mathbf{y}) / \sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} \mid \mathbf{y})$. To better understand the connection between our initial model and this extension note that, for any $v \in \mathcal{Y}$ and $\mathcal{C} \subseteq \mathcal{Y}$ such that $v \notin \mathcal{C}$, we have that

$$\frac{\eta_a(\mathcal{C} \cup \{v\} \mid \mathbf{y})}{\eta_a(\mathcal{C} \mid \mathbf{y})} = \frac{\mathrm{Q}_a(v \mid \mathbf{y})}{1 - \mathrm{Q}_a(v \mid \mathbf{y})}.$$

Thus,

$$\mathrm{Q}_a(v \mid \mathbf{y}) = \frac{\eta_a(\mathcal{C} \cup \{v\} \mid \mathbf{y})}{\eta_a(\mathcal{C} \cup \{v\} \mid \mathbf{y}) + \eta_a(\mathcal{C} \mid \mathbf{y})}.$$

Based on this alternative specification, we accommodate Assumptions A1 and A3 as follows.

(A1') For each $a \in \mathcal{A}$, $v \in \mathcal{Y}$, and $\mathbf{y} \in \overline{\mathcal{Y}}^A$, there exists $\mathcal{C} \in 2^{\mathcal{Y}}$ such that $v \succ_a v'$ for all $v' \in \mathcal{C}$ and $\eta_a(\mathcal{C} \cup \{v\} \mid \mathbf{y}) > 0$.

(A3') For each $a \in \mathcal{A}$, $\mathcal{C} \in 2^{\mathcal{Y}}$, and $\mathbf{y}, \mathbf{y}^* \in \overline{\mathcal{Y}}^A$, such that $\mathbf{y}$ is different from $\mathbf{y}^*$ just in one component a^*,

(i) $a^* \notin \mathcal{N}_a$ or $y_{a^*},\ y^*_{a^*} \notin \mathcal{C} \implies \eta_a(\mathcal{C} \mid \mathbf{y}) = \eta_a(\mathcal{C} \mid \mathbf{y}^*)$;

(ii) $a^* \in \mathcal{N}_a,\ y_{a^*} \in \mathcal{C}$ and $y^*_{a^*} \notin \mathcal{C} \implies \eta_a(\mathcal{C} \mid \mathbf{y}) > \eta_a(\mathcal{C} \mid \mathbf{y}^*)$.

Assumption A3'(i) states that the attention a person pays to a given set is invariant to the choices of those who are not connected with the person and to alternatives that do not enter the set. Assumption A3'(ii) means that switches of friends to a new alternative boosts the attention for all sets that contain this new alternative. Note that Assumption A3' does not assume that different friends affect consideration probabilities symmetrically. That is, the model allows the possibility that some friends have a bigger effect than others. Since the probability of facing consideration set $\mathcal{C}$ is proportional to the inverse of the total attention, $\sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} \mid \mathbf{y})$, even if peers are switching between alternatives that are not elements of $\mathcal{C}$, the probability of facing $\mathcal{C}$ may change.

The next proposition states that, with this general alternative specification, the network structure and the profile of preferences can be uniquely recovered from the conditional choice probabilities. Though the attention mechanism is just partially-identified, it is point identified under additional restrictions that reduce the dimensionality of the problem; we describe some of these restrictions.

[7] If one instead normalizes $\sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} \mid \mathbf{y}) = 1$, then we get the consideration rule of Aguiar (2017).

Proposition 5.3. *Suppose that Assumptions A1', A2, and A3' are satisfied. Then, the set of connections* Γ *and the profile of strict preferences* $\succ$ *are point identified from* P. *If additionally for any* $a \in \mathcal{A}$, $\mathbf{y} \in \overline{\mathcal{Y}}^{A}$, *and* $\mathcal{C}, \mathcal{D} \in 2^{\mathcal{Y}}$ *one of the following holds*

(i) (Manzini and Mariotti, 2014) $\eta_a(\mathcal{C} \mid \mathbf{y}) = \prod_{v \in \mathcal{C}} \eta_a(\{v\} \mid \mathbf{y})$;

(ii) (Dardanoni et al., 2020) $|\mathcal{C}| = |\mathcal{D}| \implies \eta_a(\mathcal{C} \mid \mathbf{y}) = \eta_a(\mathcal{D} \mid \mathbf{y})$;

(iii) $\eta_a(\mathcal{C} \mid \mathbf{y}) = \sum_{v \in \mathcal{C}} \eta_a(\{v\} \mid \mathbf{y})$;

(iv) $\eta_a(\mathcal{C} \mid \mathbf{y}) = \eta_a(\{v^*\} \mid \mathbf{y})$, *where* $v^* \in \mathcal{C}$ *satisfies* $v^* \succ_a v'$ *for all* $v' \in \mathcal{C}$;

then $\{\eta_a\}_{a \in \mathcal{A}}$ *is also pointidentified.*

Condition (i) in Proposition 5.3 restates our main result using the consideration set formation model of Manzini and Mariotti (2014). Condition (ii) shows that the model analyzed in Dardanoni et al. (2020) —where the sets of equal cardinality are considered with equal probability— also imposes enough restrictions to recover consideration probabilities. Conditions (iii) and (iv) are new. Condition (iii) is similar to condition (i), but defines "aggregate" attention as a sum of singleton-attentions. Condition (iv) postulates that consideration probability of a set only depends on the best alternative in that set.

6. Application: Choosing a Restaurant

This section simulates a sequence of choices for a simple version of our initial model that we apply to the problem of choosing a restaurant. The exercise has two aims. First, we illustrate how people's mistakes relate to the structure of the network. Second, we show how the main parts of the model can, indeed, be estimated from a sequence of choices.

6.1. Simulation

There are five people in the network. Their reference groups are as follows

$$\mathcal{N}_1 = \{2\}, \mathcal{N}_2 = \{1\}, \mathcal{N}_3 = \{1, 2\}, \mathcal{N}_4 = \{5\}, \text{ and } \mathcal{N}_5 = \{4\}.$$

Note that each person has at least one friend, so A2 is satisfied. Moreover, the network is directed since $3 \notin \mathcal{N}_1, \mathcal{N}_2$ and $1, 2 \in \mathcal{N}_3$. There are two possible restaurants at which individuals can have dinner. Restaurant 1 offers Mediterranean food and Restaurant 2 is a Steakhouse. Thus, $\mathcal{Y} = \{1, 2\}$. The default option o involves eating at home. The preferences of these people are as follows

$$2 \succ_1 1,\ 1 \succ_2 2,\ 2 \succ_3 1,\ 1 \succ_4 2, \text{ and } 1 \succ_5 2.$$

Table 1 – Marginal Equilibrium Probabilities for Directed Network

$\mu_1(o) = 0.29$	$\mu_2(o) = 0.29$	$\mu_3(o) = 0.19$	$\mu_4(o) = 0.30$	$\mu_5(o) = 0.30$
$\mu_1(1) = 0.30$	$\mu_2(1) = 0.40$	$\mu_3(1) = 0.30$	$\mu_4(1) = 0.50$	$\mu_5(1) = 0.50$
$\mu_1(2) = 0.41$	$\mu_2(2) = 0.31$	$\mu_3(2) = 0.51$	$\mu_4(2) = 0.20$	$\mu_5(2) = 0.20$

Notes: These probabilities are estimated using simulated data, sample size = 15000.

Table 2 – Probabilities of Mistakes for Directed Network

	Person 1	Person 2	Person 3	Person 4	Person 5
Probability of Mistakes	59%	60%	49%	50%	50%

Notes: These probabilities are computed using the estimated marginal equilibrium probabilities from Table 1.

That is, Persons 2, 4, and 5 prefer Mediterranean food, and Persons 1 and 3 prefer the Steakhouse. We will assume the attention mechanism is invariant across people and alternatives. To keep the exercise simple, as in Example 1, we will also assume the probability of paying attention to a given option only depends on the choices of friends. In this case, we can avoid some subindices and let $\mathrm{Q}(v \mid \mathrm{N}_a^v(\mathbf{y}))$ be the probability that Person a pays attention to restaurant $v \in \mathcal{Y}$ if $\mathrm{N}_a^v(\mathbf{y})$ people of her reference group did so last time they selected where to dine. We initially let

$$\mathrm{Q}(v \mid 0) = \frac{1}{4}, \mathrm{Q}(v \mid 1) = \frac{3}{4}, \text{ and } \mathrm{Q}(v \mid 2) = \frac{7}{8}.$$

The rates for their Poisson alarm clocks are 1.

The equilibrium behavior of this restaurant choice model is a joint distribution μ with support on 243 choice configurations (3^5). We simulated a long sequences of choices and calculated the equilibrium behavior. (See Appendix C for more details.) From the equilibrium behavior we can easily obtain the marginal equilibrium probabilities for each person. Each of them specifies the fraction of time that each person selects each alternative in the long run. These marginal probabilities are shown in Table 1.

Let us say that a person makes a mistake when her best alternative is not part of her consideration set, and thereby it is not chosen. From the previous marginals, we can easily calculate the probabilities of making mistakes. These probabilities are displayed in Table 2. Note that Persons 2 and 4 are identical in all respects except in the type of friend they have. In particular, Person 4 shares with her friend the same preferences. The opposite is true for Person 2. This difference leads Person 4 to make fewer mistakes. It becomes clear from this illustration that having friends with similar preferences helps each person to consider more often her best alternative. Thus, homophily favors people in this illustrative example. Also, note that Person 3 makes fewer mistakes since she has more friends. Thus, having more friends is also helpful in this application.

To show a bit more how the network structure shapes people's mistakes, let us add two more connections in the model. In particular, let us assume that Person 3 is a friend of Persons 1 and 2.

Table 3 – Marginal Equilibrium Probabilities for Undirected Network

$\mu_1(o) = 0.12$	$\mu_2(o) = 0.12$	$\mu_3(o) = 0.12$	$\mu_4(o) = 0.30$	$\mu_5(o) = 0.30$
$\mu_1(1) = 0.21$	$\mu_2(1) = 0.42$	$\mu_3(1) = 0.22$	$\mu_4(1) = 0.50$	$\mu_5(1) = 0.50$
$\mu_1(2) = 0.67$	$\mu_2(2) = 0.46$	$\mu_3(2) = 0.66$	$\mu_4(2) = 0.20$	$\mu_5(2) = 0.20$

Notes: These probabilities are estimated using simulated data, sample size = 15000.

Table 4 – Probabilities of Mistakes for Undirected Network

	Person 1	Person 2	Person 3	Person 4	Person 5
Probability of Mistakes	33%	58%	34%	50%	50%

Notes: These probabilities are computed using the estimated marginal equilibrium probabilities from Table 3.

That is,

$$\mathcal{N}_1 = \{2,3\}, \mathcal{N}_2 = \{1,3\}, \mathcal{N}_3 = \{1,2\}, \mathcal{N}_4 = \{5\}, \text{ and } \mathcal{N}_5 = \{4\}.$$

In this case, the network becomes undirected. Repeating the previous exercise, the new network generates the marginal equilibrium distributions depicted in Table 3.

From these marginals, the probabilities of mistakes are presented on Table 4. The probabilities of making mistakes decrease for Persons 1, 2, and 3. But the change is larger for Persons 1 and 3 as they share the same preferences over the restaurants.

6.2. Estimation

In this section we show how the observed (or simulated) data from the restaurant setup can be used to estimate the parameters of the model. Regarding the connections, we use the last specification (i.e., the undirected network). We also assume the researcher observes the sequence of choices at fixed time intervals, as in Dataset 2. Thus, identification follows by Proposition 4.3. Let us indicate by $\theta = (\Gamma, \succ, \mathrm{Q})$ an element in the space of possible parameters we want to estimate. In line with Assumption A1 and A3, attention probabilities have to be strictly between 0 and 1, and to respect the monotonicity condition. For example, in our restaurant example $\mathrm{Q}(v \mid \mathrm{N}_a^v(\mathbf{y})) \in (0,1)$ for $\mathrm{N}_a^v(\mathbf{y}) = 0, 1, 2$, and $\mathrm{Q}(v \mid 0) < \mathrm{Q}(v \mid 1) < \mathrm{Q}(v \mid 2)$.[8]

We normalize the intensity parameter λ_a to 1 for all $a \in \mathcal{A}$. Thus, for each θ, we can construct the transition rate matrix $\mathcal{M}(\theta)$ using Equations (1) and (2). In turn, this information allows to calculate the so-called transition matrix

$$\mathcal{P}(\theta, \Delta) = e^{\Delta \mathcal{M}(\theta)}.$$

We can use the latter to build the log-likelihood function $\mathrm{L}_T(\theta) = \Sigma_{t=0}^{T-1} \ln \mathcal{P}_{\iota(\mathbf{y}_t), \iota(\mathbf{y}_{t+1})}(\theta, \Delta)$, where $\iota(\mathbf{y}) \in \left\{1, 2, \ldots, \overline{\mathcal{Y}}^A\right\}$ is the position of $\mathbf{y}$ according the lexicographic order, and $\mathcal{P}_{k,m}(\theta, \Delta)$ is the

[8] Technically speaking, for estimation purposes, we can only impose weak inequalities.

Table 5 – Bias and Root Mean Squared Error (RMSE) ($\times 10^{-3}$)

		Sample Size					
Attention Probabilities		10	50	100	500	1000	5000
$Q(v \mid 0)$	Bias	213.6	164.9	133.2	65.1	42.3	9.4.
	RMSE	259.1	172.9	137.4	66.6	43.3	10.2
$Q(v \mid 1)$	Bias	66.4	53	46.9	27.2	18	2.7
	RMSE	103.9	66.9	55.6	31	20.9	5.6
$Q(v \mid 2)$	Bias	53.8	49.8	42.9	23.6	15.1	0.3
	RMSE	94.3	64.6	52.8	27.2	17.8	4.4

Notes: The sample sizes of 10, 50, 100, 500, 1000, and 5000 correspond to Δ equal to 2500, 500, 250, 50, 25, and 5. The number of replications is 1000.

(k, m)-th element of the matrix $\mathcal{P}(\theta, \Delta)$. Finally, let us define the estimated parameters as

$$\widehat{\theta}_T = \arg\max_\theta \mathrm{L}_T(\theta).$$

First, we evaluate the performance of our estimator when the network structure and preferences are known. In our Monte Carlo experiments, the estimator of the attention mechanism performs well in terms of the mean bias and the root mean squared error, as Table 5 illustrates. (See Appendix C for more details.)

Next we estimate the whole parameter vector since our method allows to consistently estimate the network structure and the preference orders of individuals as well. Since in our example, without any restrictions, there are $2^{A(A-1)} = 1,048,576$ possible networks and $(Y!)^A = 32$ strict preference orders, to make the problem computationally tractable we restricted the parameter space for Γ by making the following assumptions: (i) each person has at most two friends; (ii) the attention mechanism is invariant across people and alternatives; and (iii) the network is undirected. As a result, the number of possible networks becomes 112 (the number of possible preference orders is still 32). The experiment was conducted 500 times for different sample sizes. Table 6 presents the results of these simulations. With just 50 observations the network structure is correctly estimated 94.4 percent times. For the sample size of 500 the network structure and the preferences are correctly estimated in all simulations.

Table 6 – Correctly Estimated Network & Preferences

Sample Size	10	50	100	500
Network	32.4%	94.4%	99.8%	100%
Preferences	34.6%	85.6%	97.6%	100%
Network & Preferences	13.4%	83%	97.4%	100%

Notes: The sample sizes of 10, 50, 100, and 500 correspond to Δ equal to 2500, 500, 250, and 50. The number of replications is 500.

7. Final Remarks

This paper adds peer effects to the consideration set models. It does so by combining the dynamic model of social interactions of Blume (1993, 1995) with the (single-agent) model of random consideration sets of Manzini and Mariotti (2014). The model we build differs from most of the social interaction models, in that the choices of friends do not affect preferences but the subset of options that people end up considering. It allows us to show how the network structure can shape people's mistakes. From an applied perspective, in our model, changes in the choices of friends induce stochastic variation of the considerations sets. We show this variation can be used to recover (from a long sequence of choices) the main parts of the model. On top of nonparametrically recovering the preference ranking of each person and the attention mechanism, we identify the set of connections or edges between the people in the network. The identification strategy allows unrestricted heterogeneity across people.

References

Abaluck, J. and Adams, A. (2017). What do consumers consider before they choose? identification from asymmetric demand responses. Technical report, National Bureau of Economic Research.

Aguiar, V. H. (2017). Random categorization and bounded rationality. *Economics Letters*, 159:46–52.

Aguiar, V. H., Boccardi, M. J., and Dean, M. (2016). Satisficing and stochastic choice. *Journal of Economic Theory*, 166:445 – 482.

Aguiar, V. H., Boccardi, M. J., Kashaev, N., and Kim, J. (2019). Does random consideration explain behavior when choice is hard? evidence from a large-scale experiment.

Aguiar, V. H. and Kashaev, N. (2020). Identification and estimation of discrete outcome models with bounded covariates.

Allen, B. (1982). Some stochastic processes of interdependent demand and technological diffusion of an innovation exhibiting externalities among adopters. *International Economic Review*, 23(3):595–608.

Arnold, B. C. and Press, S. J. (1989). Compatible conditional distributions. *Journal of the American Statistical Association*, 84(405):152–156.

Barseghyan, L., Molinari, F., and Thirkettle, M. (2019). Discrete choice under risk with limited consideration. *Mimeo*.

Blevins, J. R. (2017). Identifying restrictions for finite parameter continuous time models with discrete time data. *Econometric Theory*, 33(3):739–754.

Blevins, J. R. (2018). Identification and estimation of continuous time dynamic discrete choice games. Technical report, Mimeo, Ohio State University.

Blume, L. E. (1993). The statistical mechanics of strategic interaction. *Games and economic behavior*, 5(3):387–424.

Blume, L. E. (1995). The statistical mechanics of best-response strategy revision. *Games and Economic Behavior*, 11(2):111–145.

Blume, L. E., Brock, W. A., Durlauf, S. N., and Jayaraman, R. (2015). Linear social interactions models. *Journal of Political Economy*, 123(2):444–496.

Blume, L. E. and Durlauf, S. (2003). Equilibrium concepts for social interaction models. *International Game Theory Review*, 5(3):193–209.

Bonaldi, P., Hortaçsu, A., and Kastl, J. (2015). An empirical analysis of funding costs spillovers in the euro-zone with application to systemic risk. Technical report, National Bureau of Economic Research.

Borah, A. and Kops, C. (2018). Choice via social influence. Technical report, Mimeo.

Brady, R. L. and Rehbeck, J. (2016). Menu-dependent stochastic feasibility. *Econometrica*, 84(3):1203–1223.

Brock, W. A. and Durlauf, S. N. (2001). Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235–260.

Brock, W. A. and Durlauf, S. N. (2002). A multinomial-choice model of neighborhood effects. *American Economic Review*, 92(2):298–303.

Caplin, A., Dean, M., and Leahy, J. (2018). Rational inattention, optimal consideration sets and stochastic choice.

Cattaneo, M. D., Ma, X., Masatlioglu, Y., and Suleymanov, E. (2017). A random attention model. *arXiv preprint arXiv:1712.03448*.

Chambers, C. P., Cuhadaroglu, T., and Masatlioglu, Y. (2019). A random attention model. *Mimeo*.

Conlon, C. T. and Mortimer, J. H. (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4):1–30.

Dardanoni, V., Manzini, P., Mariotti, M., and Tyson, C. J. (2020). Inferring cognitive heterogeneity from aggregate choices. *Econometrica*, 88(3):1269–1296.

De Paula, Á., Rasul, I., and Souza, P. (2018). Recovering social networks from panel data: identification, simulations and an application.

Draganska, M. and Klapper, D. (2011). Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *Journal of Marketing Research*, 48(4):653–669.

Gaynor, M., Propper, C., and Seiler, S. (2016). Free to choose? reform, choice, and consideration sets in the english national health service. *American Economic Review*, 106(11):3521–57.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica*, 76(5):1017–1074.

Horan, S. (2019). Random consideration and choice: A case study of "default" options. *Mathematical Social Sciences*, 102:73–84.

Kaiser, M. S. and Cressie, N. (2000). The construction of multivariate distributions from markov random fields. *Journal of Multivariate Analysis*, 73(2):199–220.

Lazzati, N. (2020). Co-diffusion of technologies in social networks. *American Economic Journal: Microeconomics*, Forthcoming.

Lleras, J. S., Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2017). When more is less: Limited consideration. *Journal of Economic Theory*, 170:70–85.

Manresa, E. (2013). Estimating the structure of social interactions using panel data. *Unpublished Manuscript. CEMFI, Madrid.*

Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3):229–254.

Manzini, P. and Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3):1153–1176.

Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, 102(5):2183–2205.

Mehta, N., Rajiv, S., and Srinivasan, K. (2003). Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing science*, 22(1):58–84.

Roberts, J. H. and Lattin, J. M. (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, 28(4):429–440.

A. Proofs

Proof of Proposition 3.1: For an irreducible, finite-state, continuous Markov chain the steady-state μ exists and it is unique. Thus, we only need to prove that A1 implies that the Markov chain induced by our model is irreducible. First note that, under A1, for each $a \in \mathcal{A}$, $v \in \mathcal{Y}$, and $\mathbf{y} \in \overline{\mathcal{Y}}^A$, we have that

$$1 > \mathrm{P}_a\left(v \mid \mathbf{y}\right) = \mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v\left(\mathbf{y}\right)\right) \prod_{v' \in \mathcal{Y}, v' \succ_a v} \left(1 - \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}\left(\mathbf{y}\right)\right)\right) > 0.$$

To show irreducibility, let $\mathbf{y}$ and $\mathbf{y}'$ be two different choice configurations. It follows from expression (2) that we can go from one configuration to the other one in less than A steps with positive probability. ■

Proof of Proposition 4.1: By A1, $\mathrm{P}_a\left(\cdot \mid \mathbf{y}\right)$ has full support for each $\mathbf{y} \in \overline{\mathcal{Y}}^A$. By A2 and A3, $\mathrm{P}_a\left(v \mid \mathbf{y}\right)$ is strictly decreasing in $\mathrm{N}_a^{v'}\left(\mathbf{y}\right)$ for each $v' \succ_a v$. Thus, we can recover $\mathcal{N}_a$. Since this is true for each $a \in \mathcal{A}$, we can get $\Gamma = (\mathcal{A}, e)$. Also, from variation in $\mathrm{N}_a^{v'}\left(\mathbf{y}\right)$ for each $v' \neq v$, we can recover Person a's upper level set that corresponds to option v. That is,

$$\left\{v' \in \mathcal{Y} : v' \succ_a v\right\}$$

By repeating the exercise with each alternative, we can recover $\succ_a$. Finally, suppose that y_a^* is the most preferred alternative for Person a. Then,

$$\mathrm{P}_a\left(y_a^* \mid \mathbf{y}\right) = \mathrm{Q}_a\left(y_a^* \mid y_a, \mathrm{N}_a^{y_a^*}\left(\mathbf{y}\right)\right).$$

It follows that we can recover $\mathrm{Q}_a\left(y_a^* \mid y_a, \mathrm{N}_a^{y_a^*}\left(\mathbf{y}\right)\right)$ for each $\mathbf{y} \in \overline{\mathcal{Y}}^A$. By proceeding in descending preference ordering we can then recover $\mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v\left(\mathbf{y}\right)\right)$ for all $v \in \mathcal{Y}$ (and each $\mathbf{y} \in \overline{\mathcal{Y}}^A$). ■

Proof of Proposition 4.2: Since $\lim_{\Delta \to 0} \mathcal{P}\left(\Delta\right) = \mathcal{M}$, we can recover transition rate matrix from the data. Recall that

$$\mathrm{m}\left(\mathbf{y}' \mid \mathbf{y}\right) = \begin{cases} 0 & \text{if} \quad \sum_{a \in \mathcal{A}} \mathbb{1}\left(y_a' \neq y_a\right) > 1 \\ \sum_{a \in \mathcal{A}} \lambda_a \mathrm{P}_a\left(y_a' \mid \mathbf{y}\right) \mathbb{1}\left(y_a' \neq y_a\right) & \text{if} \quad \sum_{a \in \mathcal{A}} \mathbb{1}\left(y_a' \neq y_a\right) = 1 \end{cases}.$$

Thus, $\lambda_a \mathrm{P}_a\left(y_a' \mid \mathbf{y}\right) = \mathrm{m}\left(y_a', \mathbf{y}_{-a} \mid \mathbf{y}\right)$. It follows that we can recover $\lambda_a \mathrm{P}_a\left(v \mid \mathbf{y}\right)$ for each $v \in \overline{\mathcal{Y}}$, $\mathbf{y} \in \overline{\mathcal{Y}}^A$, and $a \in \mathcal{A}$. Note that, for each $\mathbf{y} \in \overline{\mathcal{Y}}^A$,

$$\sum_{v \in \overline{\mathcal{Y}}} \lambda_a \mathrm{P}_a\left(v \mid \mathbf{y}\right) = \lambda_a \sum_{v \in \overline{\mathcal{Y}}} \mathrm{P}_a\left(v \mid \mathbf{y}\right) = \lambda_a.$$

Then we can also recover λ_a for each $a \in \mathcal{A}$. ■

Proof of Proposition 4.3: This proof builds on Theorem 1 of Blevins (2017) and Theorem 3 of Blevins (2018). For the present case, it follows from these two theorems, that the transition rate matrix $\mathcal{M}$ is generically identified if, in addition to the conditions in Proposition 4.3, we have that

$$(Y+1)^A - AY - 1 \geq \frac{1}{2}.$$

This condition is always satisfied if $A > 1$. Identification of $\mathcal{M}$ follows because, by A2, $A \geq 2$. We can then uniquely recover $(\mathrm{P}_a)_{a\in\mathcal{A}}$ from $\mathcal{M}$. See the proof of Proposition 4.2 ■

Proof of Proposition 5.1: Note that expression (3) can be rewritten as follows

$$\mathrm{P}_a(v \mid \mathbf{y}) = \sum\nolimits_{\mathcal{C}\subseteq\mathcal{Y}} \mathrm{R}_a(v \mid \mathcal{C}) \prod\nolimits_{v'\in\mathcal{C}} \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right) \prod\nolimits_{v''\notin\mathcal{C}} \left(1 - \mathrm{Q}_a\left(v'' \mid y_a, \mathrm{N}_a^{v''}(\mathbf{y})\right)\right) =$$
$$\mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^{v}(\mathbf{y})) \sum\nolimits_{\mathcal{C}\subseteq\mathcal{Y}, v\in\mathcal{C}} \mathrm{R}_a(v \mid \mathcal{C}) \prod\nolimits_{v'\in\mathcal{C}\setminus\{v\}} \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right) \prod\nolimits_{v''\notin\mathcal{C}} \left(1 - \mathrm{Q}_a\left(v'' \mid y_a, \mathrm{N}_a^{v''}(\mathbf{y})\right)\right).$$

Thus, by A2 and A3, we can state whether $a' \in \mathcal{N}_a$ by checking whether $\mathrm{P}_a(v \mid y_1 = o, \ldots, y_A = o)$ moves up when we change $y_{a'}$ from o to v for some v in $\mathcal{Y}$. It follows that the network structure is identified.

Let $\mathbf{y}$ be such that $\mathrm{N}_a^{v}(\mathbf{y}) = 0$ and let us assume that at least one person (different from a) in $\mathbf{y}$ selected the default option (i.e., there is at least one $y_{a'} = o$ with $a' \neq a$). Let $\mathbf{y}'$ be such that

$$\mathrm{N}_a^{v'}(\mathbf{y}) = \mathrm{N}_a^{v'}(\mathbf{y}') \text{ for all } v' \neq v \text{ and } \mathrm{N}_a^{v}(\mathbf{y}) = 1.$$

Note that

$$\mathrm{P}_a(v \mid \mathbf{y}') \,/\, \mathrm{P}_a(v \mid \mathbf{y}) = \mathrm{Q}_a(v \mid y_a, 1) \,/\, \mathrm{Q}_a(v \mid y_a, 0).$$

Also

$$\mathrm{P}_a(o \mid \mathbf{y}') \,/\, \mathrm{P}_a(o \mid \mathbf{y}) = (1 - \mathrm{Q}_a(v \mid y_a, 1)) \,/\, (1 - \mathrm{Q}_a(v \mid y_a, 0)).$$

Thus, by A3, $\mathrm{Q}_a(v \mid y_a, 0)$ and $\mathrm{Q}_a(v \mid y_a, 1)$ can be recovered from the data. By implementing a similar procedure for different values of $\mathrm{N}_a^{v}(\mathbf{y})$ we can recover Q_a for each y_a. Finally, since this is true for any arbitrary a and y_a, then we can recover the attention mechanism $(\mathrm{Q}_a)_{a\in\mathcal{A}}$.

We finally show that R_a is identified if and only if (in addition to A1-A3) we have that $|\mathcal{N}_a| \geq Y - 1$. We will present the idea for $v = 1$, agent a, and y_a. (The proof immediately extends to other agents and alternatives.) We want to recover $\mathrm{R}_a(1 \mid \mathcal{C})$ for all $\mathcal{C}$. To simplify the exposition we will write

$$\begin{aligned} |\mathcal{N}_a| &= N \\ \mathrm{Q}_a(v \mid y_a, m) &= \mathrm{Q}^1(v \mid m) \\ 1 - \mathrm{Q}_a(v \mid y_a, m) &= \mathrm{Q}^0(v \mid m) \end{aligned}$$

We have a set of equations indexed by $\mathbf{y}$

$$\mathrm{P}_a(1 \mid \mathbf{y}) \,/\, \mathrm{Q}_a\left(1 \mid y_a, \mathrm{N}_a^1(\mathbf{y})\right) = \sum\nolimits_{\mathcal{C}\subseteq\mathcal{Y}, v\in\mathcal{C}} \mathrm{R}_a(v \mid \mathcal{C}) \prod\nolimits_{k\in\mathcal{C}\setminus\{v\}} \mathrm{Q}^1\left(k \mid \mathrm{N}_a^k(\mathbf{y})\right) \prod\nolimits_{k\notin\mathcal{C}\setminus\{v\}} \mathrm{Q}^0\left(k \mid \mathrm{N}_a^k(\mathbf{y})\right).$$

To present the ideas more clear let $A(N, Y)$ be the matrix of coefficients in front of the $\mathrm{R}_a^{'}$s. The above system of equations has a unique solution if and only if $A(N, Y)$ has full column rank. The column of $A(N, Y)$ that corresponds to any given $\mathcal{C} \subseteq \mathcal{Y}$ consists of the elements of the following form

$$\prod\nolimits_{k \in \mathcal{Y}} \mathrm{Q}^{\mathbb{1}(k \in \mathcal{C})} \left(k \mid \mathrm{N}^k\right)$$

where $\mathrm{N}^k \in \{0, 1, \ldots, N\}$ and $\sum_k \mathrm{N}^k \leq N$. The last claim in the proposition follows from the next lemma.

Lemma 1: Assume that A1-A3 hold. For all $Y \geq 2$ and $N \geq 1$

$$N \geq Y - 1 \Longleftrightarrow A(N, Y) \text{ has full column rank.}$$

Proof.

Step 1. To illustrate how the idea works, note that $A(1, 2)$ and $A(1, 3)$ can be written as follows

$$A(1,2) = \begin{pmatrix} \mathrm{Q}^0(2 \mid 0) & \mathrm{Q}^1(2 \mid 0) \\ \mathrm{Q}^0(2 \mid 1) & \mathrm{Q}^1(2 \mid 1) \end{pmatrix}$$

and

$$\begin{aligned} A(1,3) &= \begin{pmatrix} \mathrm{Q}^0(3 \mid 0)\,\mathrm{Q}^0(2 \mid 0) & \mathrm{Q}^0(3 \mid 0)\,\mathrm{Q}^1(2 \mid 0) & \mathrm{Q}^1(3 \mid 0)\,\mathrm{Q}^0(2 \mid 0) & \mathrm{Q}^1(3 \mid 0)\,\mathrm{Q}^1(2 \mid 0) \\ \mathrm{Q}^0(3 \mid 0)\,\mathrm{Q}^0(2 \mid 1) & \mathrm{Q}^0(3 \mid 0)\,\mathrm{Q}^1(2 \mid 1) & \mathrm{Q}^1(3 \mid 0)\,\mathrm{Q}^0(2 \mid 1) & \mathrm{Q}^1(3 \mid 0)\,\mathrm{Q}^1(2 \mid 1) \\ \mathrm{Q}^0(3 \mid 1)\,\mathrm{Q}^0(2 \mid 0) & \mathrm{Q}^0(3 \mid 1)\,\mathrm{Q}^1(2 \mid 0) & \mathrm{Q}^1(3 \mid 1)\,\mathrm{Q}^0(2 \mid 0) & \mathrm{Q}^1(3 \mid 1)\,\mathrm{Q}^1(2 \mid 0) \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{Q}^0(3 \mid 0)\,A(1,2) & \mathrm{Q}^1(3 \mid 0)\,A(1,2) \\ \mathrm{Q}^0(3 \mid 1)\,A(0,2) & \mathrm{Q}^1(3 \mid 1)\,A(0,2) \end{pmatrix} \end{aligned}$$

where $A(0,2) = \begin{pmatrix} \mathrm{Q}^0(2 \mid 0) & \mathrm{Q}^1(2 \mid 0) \end{pmatrix}$.

Similarly, the matrix $A(N, Y+1)$ can be written as follows

$$A(N, Y+1) = \begin{pmatrix} \mathrm{Q}^0(Y+1 \mid 0)\,A(N,Y) & \mathrm{Q}^1(Y+1 \mid 0)\,A(N,Y) \\ \mathrm{Q}^0(Y+1 \mid 1)\,A(N-1,Y) & \mathrm{Q}^1(Y+1 \mid 1)\,A(N-1,Y) \\ \mathrm{Q}^0(Y+1 \mid 2)\,A(N-2,Y) & \mathrm{Q}^1(Y+1 \mid 2)\,A(N-2,Y) \\ \cdots & \cdots \\ \mathrm{Q}^0(Y+1 \mid N)\,A(0,Y) & \mathrm{Q}^1(Y+1 \mid N)\,A(0,Y) \end{pmatrix}.$$

Note that $A(K, Y)$ is a submatrix of $A(K+1, Y)$ for all K (with the same number of columns). Thus, it is clear that $A(N, Y+1)$ has full column rank only if $A(N, Y)$ and $A(N-1, Y)$ have both full column rank, which is the same as to say $A(N-1, Y)$ has full column rank. We next show that under A3, if $A(N-1, Y)$ has full column rank, then $A(N, Y+1)$ has full column rank too. To this end, let M be a matrix obtained deleting rows from $A(N-1, Y)$ in such a way that $\det(M) > 0$.

Then, by A3, we have

$$\det\begin{pmatrix} \mathrm{Q}^0\left(Y+1 \mid 0\right)M & \mathrm{Q}^1\left(Y+1 \mid 0\right)M \\ \mathrm{Q}^0\left(Y+1 \mid 1\right)M & \mathrm{Q}^1\left(Y+1 \mid 1\right)M \end{pmatrix} = \left(\mathrm{Q}^1\left(Y+1 \mid 1\right) - \mathrm{Q}^1\left(Y+1 \mid 0\right)\right)^{2^{Y-1}} \det\left(M\right)^2 > 0.$$

In summary, we have that

$$A(N, Y+1) \text{ has full column rank } \iff A(N-1, Y) \text{ has full column rank.}$$

Step 2. Consider $(N, Y) = (1, 2)$. Note that

$$A(1,2) = \begin{pmatrix} \mathrm{Q}^0\left(2 \mid 0\right) & \mathrm{Q}^1\left(2 \mid 0\right) \\ \mathrm{Q}^0\left(2 \mid 1\right) & \mathrm{Q}^1\left(2 \mid 1\right) \end{pmatrix}$$

has full column rank since $\det\left(A(1,2)\right) = \mathrm{Q}^1\left(2 \mid 1\right) - \mathrm{Q}^1\left(2 \mid 0\right) > 0$. Also, any $A(N, 2)$ with $N \geq 1$ will have full column rank because $A(1,2)$ is a submatrix of $A(N,2)$ with the same number of columns.

Finally, note that $A(1,3)$ does not have full column rank since the number of columns is higher than the number of rows.

Step 3. From Step 1 we got that, for all $Y \geq 2$ and $N \geq 1$,

$$A(N, Y+1) \text{ has full column rank } \iff A(N-1, Y) \text{ has full column rank.}$$

From Step 2, we get that $A(N,2)$ (with $N \geq 1$) has full column rank and $A(1,3)$ does not have full column rank. The claim in Lemma 1 follows by combining these three results. ■

Proof of Proposition 5.2: This proof is divided in three steps.

Step 1. (Identification of the Set of Connections) Take any two different people with arbitrary designations a_1 and a_2. Note that if $a_2 \notin \mathcal{N}_{a_1}$, then $\mathrm{P}_{a_1}(v \mid \mathbf{y}) = \mathrm{P}_{a_1}(v \mid \mathbf{y}')$ for any $\mathbf{y}$ and $\mathbf{y}'$ such that $y_a = y'_a$ for all $a \neq a_2$ and $y_{a_1} \neq v$. Also, let $v^*_{a_1}$ be the best preferred alternative of a_1. Then, by A1 and A3, for any $\mathbf{y}$ such that $y_{a_1} \neq v^*_{a_1}$

$$\mathrm{P}_{a_1}\left(v^*_{a_1} \mid \mathbf{y}\right) = \mathrm{Q}_{a_1}\left(v^*_{a_1} \mid \mathbf{y}\right)$$

is constant in y_{a_2} if and only if $a_2 \notin \mathcal{N}_{a_1}$. Altogether, $a_2 \notin \mathcal{N}_{a_1}$ if and only if $\mathrm{P}_{a_1}\left(v \mid \mathbf{y}\right)$ with $y_{a_1} \neq v$ is constant in y_{a_2}. As a result, we can identify whether a_2 is in the set of friends of a_1. Since the choice of a_1 and a_2 was arbitrary, we can identify the set of connections Γ. Note that for this result we only need $|\mathcal{Y}| \geq 2$.

Step 2. (Identification of the Preferences) Fix some Person a_1. We will show the result for a set of alternatives $\mathcal{Y} = \{1, 2, 3\}$ of size 3. (The proof easily extends to the case of more alternatives.

The only cost is extra notation.) For arbitrary designation of the three options v_1, v_2, v_3, and any $a_2 \in \mathcal{N}_{a_1}$ (by A2, $\mathcal{N}_{a_1} \neq \emptyset$) let

$$\Delta \mathrm{P}(v_1, v_2, v_3) = \mathrm{P}_{a_1}(v_1 \mid (v_3, \ldots, v_3, v_2, v_3, \ldots, v_3)) - \mathrm{P}_{a_1}(v_1 \mid (v_3, \ldots, v_3, v_3, v_3, \ldots, v_3)),$$

where Person a_2 switches from v_2 to v_3. Define $\mathrm{sign}(v_1, v_2, v_3) \in \{-, +, 0\}$ be such that

$$\mathrm{sign}(v_1, v_2, v_3) = \begin{cases} - & \text{if } \Delta \mathrm{P}(v_1, v_2, v_3) < 0 \\ + & \text{if } \Delta \mathrm{P}(v_1, v_2, v_3) > 0 \\ 0 & \text{if } \Delta \mathrm{P}(v_1, v_2, v_3) = 0 \end{cases}.$$

Note that $\mathrm{sign}(\cdot)$ can be computed from the data. Different preference orders over $\mathcal{Y}$ will imply potentially different values for $\mathrm{sign}(\cdot)$. Let N_{a_1} is the the cardinality of $\mathcal{N}_{a_1}$. For example, if $1 \succ_{a_1} 2 \succ_{a_1} 3$, then we have that

$$\begin{aligned}
\Delta \mathrm{P}(1,2,3) &= \mathrm{Q}_{a_1}(1 \mid 3, 0) - \mathrm{Q}_{a_1}(1 \mid 3, 0) = 0 \\
\Delta \mathrm{P}(1,3,2) &= \mathrm{Q}_{a_1}(1 \mid 2, 0) - \mathrm{Q}_{a_1}(1 \mid 3, 0) = 0 \\
\Delta \mathrm{P}(3,1,2) &= \mathrm{Q}_{a_1}(3 \mid 2, 0)(1 - \mathrm{Q}_{a_1}(1 \mid 2, 1))(1 - \mathrm{Q}_{a_1}(2 \mid 2, \mathrm{N}_{a_1} - 1)) \\
&\quad - \mathrm{Q}_{a_1}(3 \mid 2, 0)(1 - \mathrm{Q}_{a_1}(1 \mid 2, 0))(1 - \mathrm{Q}_{a_1}(2 \mid 2, \mathrm{N}_{a_1})) \\
\Delta \mathrm{P}(3,2,1) &= \mathrm{Q}_{a_1}(3 \mid 1, 0)(1 - \mathrm{Q}_{a_1}(1 \mid 1, \mathrm{N}_{a_1} - 1))(1 - \mathrm{Q}_{a_1}(2 \mid 1, 1)) \\
&\quad - \mathrm{Q}_{a_1}(3 \mid 1, 0)(1 - \mathrm{Q}_{a_1}(1 \mid 1, \mathrm{N}_{a_1}))(1 - \mathrm{Q}_{a_1}(2 \mid 1, 0)),
\end{aligned}$$

Note that for $\Delta \mathrm{P}(3,1,2)$ and $\Delta \mathrm{P}(3,2,1)$ our monotonicity restriction is not informative enough: $\Delta \mathrm{P}(3,1,2)$ and $\Delta \mathrm{P}(3,2,1)$ can be positive, negative, or equal to zero. The following table displays the values $\mathrm{sign}(\cdot)$ takes depending on the underlying preference order $\succ_{a_1}$ for all distinct v_1, v_2, and v_3. If the sign of $\Delta \mathrm{P}(v_1, v_2, v_3)$ is not uniquely determined by a strict preference, then we write $\sim$.

order/(v_1, v_2, v_3)	$(1,2,3)$	$(1,3,2)$	$(2,1,3)$	$(2,3,1)$	$(3,1,2)$	$(3,2,1)$
$1 \succ_{a_1} 2 \succ_{a_1} 3$	0	0	$-$	$+$	$\sim$	$\sim$
$1 \succ_{a_1} 3 \succ_{a_1} 2$	0	0	$\sim$	$\sim$	$-$	$+$
$2 \succ_{a_1} 1 \succ_{a_1} 3$	$-$	$+$	0	0	$\sim$	$\sim$
$2 \succ_{a_1} 3 \succ_{a_1} 1$	$\sim$	$\sim$	0	0	$+$	$-$
$3 \succ_{a_1} 1 \succ_{a_1} 2$	$+$	$-$	$\sim$	$\sim$	0	0
$3 \succ_{a_1} 2 \succ_{a_1} 1$	$\sim$	$\sim$	$+$	$-$	0	0

Note that we can always distinguish $1 \succ_{a_1} 2 \succ_{a_1} 3$ from say $3 \succ_{a_1} 2 \succ_{a_1} 1$ since $\mathrm{sign}(2,1,3)$ is determined for both preference orders and gives different predictions. The only pairs of preference orders that may be observationally equivalent are the following three pairs: (i) $1 \succ_{a_1} 2 \succ_{a_1} 3$ and $1 \succ_{a_1} 3 \succ_{a_1} 2$; (ii) $2 \succ_{a_1} 1 \succ_{a_1} 3$ and $2 \succ_{a_1} 3 \succ_{a_1} 1$; and (iii) $3 \succ_{a_1} 1 \succ_{a_1} 2$ and $3 \succ_{a_1} 2 \succ_{a_1} 1$. Although we cannot uniquely identify the order, we can uniquely identify the most preferred alternative. For example, if we know that $1 \succ_{a_1} 2 \succ_{a_1} 3$ or $1 \succ_{a_1} 2 \succ_{a_1} 3$ has generated the data, then 1 is the most preferred alternative for agent a_1.

Assume, without loss of generality, that 3 is the most preferred alternative. Then we can identify $\mathrm{Q}_{a_1}(3 \mid y_{a_1}, \cdot)$ for any $y_{a_1} \neq 3$ since

$$\mathrm{Q}_{a_1}\left(3 \mid y_{a_1}, \mathrm{N}_{a_1}^3(\mathbf{y})\right) = \mathrm{P}_{a_1}(3 \mid \mathbf{y}).$$

Hence, for $\mathbf{y}$ with $y_{a_1} = 1$ we have that

$$\frac{\mathrm{P}_{a_1}(2 \mid \mathbf{y})}{1 - \mathrm{P}_{a_1}(3 \mid \mathbf{y})} = \begin{cases} \mathrm{Q}_a(2 \mid 1, \mathrm{N}_{a_1}^2(\mathbf{y})), & \text{if } 2 \succ_{a_1} 1 \\ \mathrm{Q}_a(2 \mid 1, \mathrm{N}_{a_1}^2(\mathbf{y}))(1 - \mathrm{Q}_a(1 \mid 1, \mathrm{N}_{a_1}^1(\mathbf{y}))), & \text{if } 1 \succ_{a_1} 2. \end{cases}$$

Start with a $\mathbf{y}$ such that $y_a = 3$ for all $a \neq a_1$ (and $y_{a_1} = 1$). Consider then changing the y_{a_2} from 3 to 1 for some $a_2 \in \mathcal{N}_{a_1}$. Then, by A3, $1 \succ_{a_1} 2$ if and only if $\mathrm{P}_{a_1}(2 \mid \mathbf{y}) / (1 - \mathrm{P}_{a_1}(3 \mid \mathbf{y}))$ strictly decreases in the data.

Step 3. (Identification of the Attention Mechanism) Fix some a_1 and let $y^*_{a_1}$ be the most preferred alternative of Person a_1. Then we can identify $\mathrm{Q}_{a_1}(y^*_{a_1} \mid y_{a_1}, \mathrm{N}_{a_1}^{y^*_{a_1}}(\mathbf{y}))$ for any $\mathbf{y}$ such that $y^*_{a_1} \neq y_{a_1}$ since

$$\mathrm{Q}_{a_1}\left(y^*_{a_1} \mid y_{a_1}, \mathrm{N}_{a_1}^{y^*_{a_1}}(\mathbf{y})\right) = \mathrm{P}_{a_1}(y^*_{a_1} \mid \mathbf{y}).$$

By proceeding in decreasing preference order we can recover $\mathrm{Q}_{a_1}(y'_{a_1} \mid y_{a_1}, N_{a_1}^{y'_{a_1}}(\mathbf{y}))$ for any y'_{a_1} and $\mathbf{y}$ such that $y'_{a_1} \neq y_{a_1}$. Moreover, we can identify

$$\prod\nolimits_{v' \neq y_{a_1}} \left(1 - \mathrm{Q}_a\left(v' \mid y_{a_1}, \mathrm{N}_{a_1}^{v'}(\mathbf{y})\right)\right)$$

Next note that for any $\mathbf{y}$ such that $y^*_{a_1} = y_{a_1}$

$$\mathrm{Q}_{a_1}(y^*_{a_1} \mid y_{a_1}, \mathrm{N}_{a_1}^{y^*_{a_1}}(\mathbf{y})) = \frac{\mathrm{P}_{a_1}(y^*_{a_1} \mid \mathbf{y}) - \prod_{v' \neq y^*_{a_1}} \left(1 - \mathrm{Q}_a\left(v' \mid y^*_{a_1}, \mathrm{N}_{a_1}^{v'}(\mathbf{y})\right)\right)}{1 - \prod_{v' \neq y^*_{a_1}} \left(1 - \mathrm{Q}_a\left(v' \mid y^*_{a_1}, \mathrm{N}_{a_1}^{v'}(\mathbf{y})\right)\right)}.$$

Hence, we can identify $\mathrm{Q}_{a_1}(y^*_{a_1} \mid y_{a_1}, \mathrm{N}_{a_1}^{y^*_{a_1}}(\mathbf{y}))$ for all $\mathbf{y}$. Let $y^{**}_{a_1}$ be the second best alternative of Person a_1, then for any $\mathbf{y}$ such that $y^{**}_{a_1} = y_{a_1}$ similarly to the case with $y^*_{a_1}$ we can identify $\mathrm{Q}_{a_1}(y^{**}_{a_1} \mid y_{a_1}, \mathrm{N}_{a_1}^{y^{**}_{a_1}}(\mathbf{y}))$ since

$$\mathrm{Q}_{a_1}(y^{**}_{a_1} \mid y_{a_1}, N_{a_1}^{y^{**}_{a_1}}(\mathbf{y})) = \frac{\mathrm{P}_{a_1}(y^{**}_{a_1} \mid \mathbf{y}) - \prod_{v' \neq y^{**}_{a_1}} \left(1 - \mathrm{Q}_{a_1}\left(v' \mid y^{**}_{a_1}, \mathrm{N}_{a_1}^{v'}(\mathbf{y})\right)\right)}{1 - \mathrm{Q}_{a_1}(y^*_{a_1} \mid y_{a_1}, \mathrm{N}_{a_1}^{v'}(\mathbf{y})) - \prod_{v' \neq y^{**}_{a_1}} \left(1 - \mathrm{Q}_{a_1}\left(v' \mid y^{**}_{a_1}, \mathrm{N}_{a_1}^{v'}(\mathbf{y})\right)\right)},$$

and thus we recover $\mathrm{Q}_{a_1}(y^{**}_{a_1} \mid y_{a_1}, N_{a_1}^{y^{**}_{a_1}}(\mathbf{y}))$ for all $\mathbf{y}$. By proceeding in decreasing preference order we can recover $\mathrm{Q}_{a_1}(y'_{a_1} \mid y_{a_1}, N_{a_1}^{y'_{a_1}}(\mathbf{y}))$ for any y'_{a_1} and $\mathbf{y}$. Since the choice of a_1 was arbitrary we can identify $(\mathrm{Q}_a)_{a \in \mathcal{A}}$.

Proof of Proposition 5.3: Step 1. (Identification of the Set of Connections) Take any two different agents a_1 and a_2. Note that if $a_2 \notin \mathcal{N}_{a_1}$, then $\mathrm{P}_{a_1}(v \mid \mathbf{y}) = \mathrm{P}_{a_1}(v \mid \mathbf{y}')$ for any $\mathbf{y}$ and $\mathbf{y}'$ such that $y_a = y'_a$ for all $a \neq a_2$ and $y_{a_1} \neq v$. Also, if $v^*_{a_1}$ is the best preferred alternative of a_1, then by Assumptions A1' and A3', for any $\mathbf{y}$

$$\frac{\mathrm{P}_{a_1}\left(v^*_{a_1} \mid \mathbf{y}\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}\right)} = \sum\nolimits_{\mathcal{C} \in 2^{\mathcal{Y}}:\, v^*_{a_1} \in \mathcal{C}} \eta_{a_1}\left(\mathcal{C} \mid \mathbf{y}\right)$$

is constant in choices of Person a_2 if and only if $a_2 \notin \mathcal{N}_{a_1}$. Hence, $a_2 \notin \mathcal{N}_{a_1}$ if and only if $\mathrm{P}_{a_1}\left(v \mid \mathbf{y}\right) / \mathrm{P}_{a_1}\left(o \mid \mathbf{y}\right)$ is constant in y_{a_2}. As a result, we can identify whether a_2 is a friend of a_1. Since the choice of a_1 and a_2 was arbitrary we can identify the whole Γ. Note that for this result to hold we only need $|\mathcal{Y}| \geq 2$.

Step 2. (Identification of the Preferences) Fix Person a_1 and $\mathbf{y}^* = (o, o, \ldots, o)\prime$. Note that

$$\frac{\mathrm{P}_{a_1}\left(v^*_{a_1} \mid \mathbf{y}^*\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}^*\right)} = \sum\nolimits_{\mathcal{C} \in 2^{\mathcal{Y}}:\, v^*_{a_1} \in \mathcal{C}} \eta_{a_1}\left(\mathcal{C} \mid \mathbf{y}^*\right)$$

will increase if any of friends of a_1 switches to anything else. Let $v^{**}_{a_1}$ be the second best preferred alternative of a_1. Then

$$\frac{\mathrm{P}_{a_1}\left(v^{**}_{a_1} \mid \mathbf{y}^*\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}^*\right)} = \sum\nolimits_{\mathcal{C} \in 2^{\mathcal{Y}}:\, v^*_{a_1} \notin \mathcal{C},\, v^{**}_{a_1} \in \mathcal{C}} \eta_{a_1}\left(\mathcal{C} \mid \mathbf{y}^*\right)$$

will increase if any of friends of a_1 switches to anything else but $v^*_{a_1}$. Moreover, this probability will not change if any of friends of a_1 switches to $v^*_{a_1}$. Hence, we can identify $v^*_{a_1}$. Applying the above step in decreasing order, we can identify the whole preference order of Person a_1. Since the choice of a_1 was arbitrary we identify preferences of all persons.

Step 3. (Identification of the Attention Mechanism) Take any Person a_1 and configuration $\mathbf{y}$. Since preferences are identified, assume without loss of generality that $Y \succ_{a_1} Y-1 \succ_{a_1} Y-2 \succ_{a_1} \cdots \succ_{a_1} 2 \succ_{a_1} 1$. Note that we have the following system of Y equations

$$\frac{\mathrm{P}_{a_1}\left(k \mid \mathbf{y}\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}\right)} = \begin{cases} \eta_{a_1}(\{1\} \mid \mathbf{y}), & k = 1, \\ \sum_{\mathcal{C} \subseteq \{1, \ldots, k-1\}} \eta_{a_1}(\mathcal{C} \cup \{k\} \mid \mathbf{y}), & k = 2, \ldots, Y. \end{cases}$$

Note that η_{a_1} could have generated the data if and only if it solves the above system of equations. Since, there are $2^Y - 1$ unknown parameters (recall that attention to the empty set is normalized to be 1) and Y equations, and there is no single attention parameter that enters more than one equation, η_{a_1} can not be identified without more restrictions. Suppose η_{a_1} is multiplicative. So we only need to identify $\eta_{a_1}(\{k\} \mid \mathbf{y})$ for all $k \in \mathcal{Y}$. Then, first, we can identify $\eta_{a_1}(\{1\} \mid \mathbf{y})$ since

$$\frac{\mathrm{P}_{a_1}\left(1 \mid \mathbf{y}\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}\right)} = \eta_{a_1}(\{1\} \mid \mathbf{y}).$$

Next,

$$\frac{\mathrm{P}_{a_1}(2 \mid \mathbf{y})}{\mathrm{P}_{a_1}(o \mid \mathbf{y})} = \eta_{a_1}(\{2\} \mid \mathbf{y}) + \eta_{a_1}(\{1,2\} \mid \mathbf{y}) = \eta_{a_1}(\{2\} \mid \mathbf{y}) + \eta_{a_1}(\{1\} \mid \mathbf{y})\eta_{a_1}(\{2\} \mid \mathbf{y}).$$

Hence, we identify $\eta_{a_1}(\{2\} \mid \mathbf{y})$. Repeating the above steps recursively we can identify η_{a_1} since

$$\eta_{a_1}(\{k\} \mid \mathbf{y}) = \frac{\mathrm{P}_{a_1}(k \mid \mathbf{y})}{\mathrm{P}_{a_1}(o \mid \mathbf{y})} \cdot \frac{1}{\sum_{\mathcal{C} \subseteq \{1,\ldots,k-1\}} \eta_a(\mathcal{C} \mid \mathbf{y})}.$$

The same argument can be applied if η_{a_1} is additive. This approach can be generalized if one models the attention index of a set as a strictly increasing transformation of attention indexes of elements of that set (i.e., $\eta_{a_1}(\{1,2\} \mid \mathbf{y}) = \phi_{\{1,2\}}(\eta_{a_1}(\{1\} \mid \mathbf{y}), \eta_{a_1}(\{2\} \mid \mathbf{y}))$, where $\phi_{1,2}$ is strictly increasing in both arguments).

Suppose that η_{a_1} is the same for sets of the same cardinality. Since we know $\eta_{a_1}(\{1\})$ from the first equation we identify attention indexes for all singleton sets including $\eta_{a_1}(\{2\} \mid \mathbf{y})$. Hence, using the second equation we identify $\eta_{a_1}(\{1,2\} \mid \mathbf{y})$ and, thus, attention indexes for all sets of cardinality 2. Repeating the above arguments recursively we can identify η_{a_1}.

It is left to show that if η_{a_1} is the same for sets that have the same best option, then it is also identified. Again, $\eta_{a_1}(\{1\})$is identified from the first equation. Since $\eta_{a_1}(\{2\} \mid \mathbf{y}) = \eta_{a_1}(\{1,2\} \mid \mathbf{y})$, from the second equation we identify

$$\eta_{a_1}(\{2\} \mid \mathbf{y}) = \eta_{a_1}(\{1,2\} \mid \mathbf{y}) = \frac{\mathrm{P}_{a_1}(2 \mid \mathbf{y})}{2\,\mathrm{P}_{a_1}(o \mid \mathbf{y})}.$$

Repeating the above arguments for every equation we get that

$$\eta_{a_1}(\mathcal{C} \mid \mathbf{y}) = \frac{\mathrm{P}_{a_1}\left(v^*_{a_1,\mathcal{C}} \mid \mathbf{y}\right)}{v^*_{a_1,\mathcal{C}}\,\mathrm{P}_{a_1}(o \mid \mathbf{y})},$$

where $v^*_{a_1,\mathcal{C}}$ is the best alternative in $\mathcal{C}$ according to $\succ_{a_1}$. Since the choice of a_1 and $\mathbf{y}$ was arbitrary, we identify η_a for all $a \in \mathcal{A}$.

We conclude the proof by showing that restrictions (i) and (ii) correspond to the models of consideration sets formation in Manzini and Mariotti (2014) and Dardanoni et al. (2020), respectively. Equivalence between (ii) and the model in Dardanoni et al. (2020) is straightforward since

$$\frac{\eta_a(\mathcal{C} \mid \mathbf{y})}{\sum_{\mathcal{B} \subseteq \mathcal{Y}} \eta_a(\mathcal{B} \mid \mathbf{y})} = \frac{\eta_a(\mathcal{D} \mid \mathbf{y})}{\sum_{\mathcal{B} \subseteq \mathcal{Y}} \eta_a(\mathcal{B} \mid \mathbf{y})} \iff \eta_a(\mathcal{C} \mid \mathbf{y}) = \eta_a(\mathcal{D} \mid \mathbf{y}).$$

To show equivalence between (i) and the model in Manzini and Mariotti (2014) assume first that $\eta_a(\cdot \mid \mathbf{y})$ is multiplicative (i.e., satisfies condition (i)). Then for every $v \in \mathcal{Y}$ define

$$\mathrm{Q}_a(v \mid \mathbf{y}) = \frac{\eta_a(\{y\} \mid \mathbf{y})}{1 + \eta_a(\{v\} \mid \mathbf{y})}.$$

Since $\eta_a(\{v\} \mid \mathbf{y}) > 0$ (otherwise multiplicativity of η_a would imply that sets that contain v are never considered, which would contradict Assumption A1') and finite, we get that $\mathrm{Q}_a(v \mid \mathbf{y})) \in (0,1)$. Take any A. By multiplicativity of η_a we get that

$$\eta_a(A \mid \mathbf{y}) = \prod\nolimits_{y \in A} \eta_a(\{y\} \mid \mathbf{y}).$$

Note that for singleton $\mathcal{Y}$ we have that $\sum_{C \subseteq \mathcal{Y}} \eta_a(C \mid \mathbf{y}) = 1 + \eta_a(\mathcal{Y} \mid \mathbf{y})$. Then using multiplicativity we get that for $\mathcal{Y} \cup \{y^*\}$ such that $y^* \notin \mathcal{Y}$

$$\sum\nolimits_{C \subseteq \mathcal{Y} \cup \{y^*\}} \eta_a(C \mid \mathbf{y}) = \sum\nolimits_{C \subseteq \mathcal{Y}} \eta_a(C \mid \mathbf{y}) + \sum\nolimits_{C \subseteq \mathcal{Y}} \eta_a(C \cup \{y^*\} \mid \mathbf{y}) = (1 + \eta_a(\{y^*\} \mid \mathbf{y})) \sum\nolimits_{C \subseteq \mathcal{Y}} \eta_a(C \mid \mathbf{y}).$$

Hence, by induction

$$\sum\nolimits_{C \subseteq \mathcal{Y}} \eta_a(C \mid \mathbf{y}) = \prod\nolimits_{y \in \mathcal{Y}} (1 + \eta_a(\{y\} \mid \mathbf{y})).$$

As a result,

$$\begin{aligned} \frac{\eta_a(A \mid \mathbf{y})}{\sum_{C \subseteq \mathcal{Y}} \eta_a(C \mid \mathbf{y})} &= \prod\nolimits_{y \in A} \frac{\eta_a(\{y\} \mid \mathbf{y})}{1 + \eta_a(\{y\} \mid \mathbf{y})} \prod\nolimits_{y' \in \mathcal{Y} \setminus A} \left(1 - \frac{\eta_a(\{y'\} \mid \mathbf{y})}{1 + \eta_a(\{y'\} \mid \mathbf{y})}\right) \\ &= \prod\nolimits_{y \in A} \mathrm{Q}_a(y \mid \mathbf{y}) \prod\nolimits_{y' \in \mathcal{Y} \setminus A} (1 - \mathrm{Q}_a(y' \mid \mathbf{y})). \end{aligned}$$

Thus, condition (i) implies the model in Manzini and Mariotti (2014).

To show the opposite assume that the probability of considering a set $\mathcal{A}$ is

$$\prod\nolimits_{y \in A} \mathrm{Q}_a(y \mid \mathbf{y}) \prod\nolimits_{y' \in \mathcal{Y} \setminus A} (1 - \mathrm{Q}_a(y' \mid \mathbf{y}))$$

for some Q_a. For singleton sets define $\eta_a(\{y\} \mid \mathbf{y}) = \dfrac{\mathrm{Q}_a(y \mid \mathbf{y})}{1 - \mathrm{Q}_a(y \mid \mathbf{y})}$. For sets with cardinality bigger than one define $\eta_a(A \mid \mathbf{y}) = \prod_{y \in A} \eta_a(\{y\} \mid \mathbf{y})$. By construction, the constructed η_a satisfies multiplicativity and it is easy to verify that

$$\prod\nolimits_{y \in A} \mathrm{Q}_a(y \mid \mathbf{y}) \prod\nolimits_{y' \in \mathcal{Y} \setminus A} (1 - \mathrm{Q}_a(y' \mid \mathbf{y})) = \frac{\eta_a(A \mid \mathbf{y})}{\sum_{C \subseteq \mathcal{Y}} \eta_a(C \mid \mathbf{y})}.$$

B. Gibbs Random Field Model

The starting point of the Gibbs random field models is a set of conditional probability distributions. In our model, the set of conditional probabilities is $(\mathrm{P}_a)_{a \in \mathcal{A}}$, with a generic element given by

$$\mathrm{P}_a(v \mid \mathbf{y}) = \mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})\right) \prod\nolimits_{v' \in \mathcal{Y}, v' \succ_a v} \left(1 - \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right)\right) \text{ for } v \in \mathcal{Y}.$$

A Gibbs equilibrium is defined as a joint distribution over the vector of choices $\mathbf{y}$, $\mathrm{P}(\mathbf{y})$, that is able to generate $(\mathrm{P}_a)_{a\in\mathcal{A}}$ as its conditional distribution functions.

Gibbs equilibria typically do not exist. (In the statistical literature, a similar existence problem is referred as the issue of compatibility of conditional distributions.) The existence of Gibbs equilibria depends on a great deal of homogeneity among people. In our model, it would also require $\mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y}))$ to be invariant with respect to y_a. Condition G1 captures this restriction.

(G1) For each $a \in \mathcal{A}$, $v \in \mathcal{Y}$, and $\mathbf{y} \in \overline{\mathcal{Y}}^A$, $\mathrm{Q}_a(v \mid \mathbf{y}) \equiv \mathrm{Q}_a(v \mid \mathrm{N}_a^v(\mathbf{y}))$.

Together with Assumption A1, this extra condition allows a simple characterization of the invariant distribution μ. We describe this characterization next.

Proposition B.1. *If A1 and G1 are satisfied, then there exists a unique μ. Also, μ satisfies*

$$\mu(\mathbf{y}) = \frac{1}{\sum_{a\in\mathcal{A}}\lambda_a}\sum\nolimits_{a\in\mathcal{A}}\lambda_a \mathrm{P}_a(y_a \mid \mathbf{y})\,\mu_{-a}(\mathbf{y}_{-a}) \text{ for each } \mathbf{y} \in\overline{\mathcal{Y}}^A.$$

Proof of Proposition B.1: The characterization of μ follows as the invariant distribution satisfies the balance condition $\sum_{\mathbf{y}'\in\overline{\mathcal{Y}}^A}\mu(\mathbf{y}')\,\mathrm{m}(\mathbf{y} \mid \mathbf{y}') = 0$ for each $\mathbf{y} \in\overline{\mathcal{Y}}^A$. The next steps show this claim.

$$\begin{aligned}
\sum\nolimits_{\mathbf{y}'\in\overline{\mathcal{Y}}^A}\mu(\mathbf{y}')\,\mathrm{m}(\mathbf{y}\mid\mathbf{y}') &= 0\\
\mu(\mathbf{y})\left(-\sum\nolimits_{\mathbf{y}'\in\overline{\mathcal{Y}}^A\setminus\{\mathbf{y}\}}\mathrm{m}(\mathbf{y}'\mid\mathbf{y})\right)+\sum\nolimits_{\mathbf{y}'\in\overline{\mathcal{Y}}^A\setminus\{\mathbf{y}\}}\mu(\mathbf{y}')\,\mathrm{m}(\mathbf{y}\mid\mathbf{y}') &= 0\\
-\mu(\mathbf{y})\sum\nolimits_{a\in\mathcal{A}}\sum\nolimits_{y_a'\in\overline{\mathcal{Y}}\setminus\{y_a\}}\lambda_a\mathrm{P}_a(y_a'\mid\mathbf{y})+\sum\nolimits_{a\in\mathcal{A}}\sum\nolimits_{y_a'\in\overline{\mathcal{Y}}\setminus\{y_a\}}\mu(y_a',\mathbf{y}_{-a})\,\lambda_a\mathrm{P}_a(y_a\mid y_a',\mathbf{y}_{-a}) &= 0\\
-\mu(\mathbf{y})\sum\nolimits_{a\in\mathcal{A}}\lambda_a(1-\mathrm{P}_a(y_a\mid\mathbf{y}))+\sum\nolimits_{a\in\mathcal{A}}\sum\nolimits_{y_a'\in\overline{\mathcal{Y}}\setminus\{y_a\}}\mu(y_a',\mathbf{y}_{-a})\,\lambda_a\mathrm{P}_a(y_a\mid y_a',\mathbf{y}_{-a}) &= 0\\
\frac{1}{\sum_{a\in\mathcal{A}}\lambda_a}\sum\nolimits_{a\in\mathcal{A}}\lambda_a\left\{\sum\nolimits_{y_a'\in\overline{\mathcal{Y}}}\mu(y_a',\mathbf{y}_{-a})\,\mathrm{P}_a(y_a\mid y_a',\mathbf{y}_{-a})\right\} &= \mu(\mathbf{y})\\
\frac{1}{\sum_{a\in\mathcal{A}}\lambda_a}\sum\nolimits_{a\in\mathcal{A}}\lambda_a\mathrm{P}_a(y_a\mid\mathbf{y})\,\mu_{-a}(\mathbf{y}_{-a}) &= \mu(\mathbf{y}).
\end{aligned}$$

In moving from the fifth line to the sixth one we used the fact that, in our model, $\mathrm{P}_a(y_a \mid y_a', \mathbf{y}_{-a}) = \mathrm{P}_a(y_a \mid \mathbf{y}_{-a})$ for any $y_a'\in\overline{\mathcal{Y}}^A$. ■

In what follows, we will normalize the lambdas to the value of 1. From Proposition B.1, μ satisfies

$$\mu(\mathbf{y}) = \frac{1}{A}\sum\nolimits_{a\in\mathcal{A}}\mathrm{P}_a(y_a \mid \mathbf{y})\,\mu_{-a}(\mathbf{y}_{-a}) \text{ for each } \mathbf{y} \in\overline{\mathcal{Y}}^A. \tag{5}$$

We next show that if $(\mathrm{P}_a)_{a\in A}$ is a set of compatible conditional distributions, then $\mu = \mathrm{P}$ solves (5). If we let $\mu = \mathrm{P}$, then right hand side of (5) is

$$\frac{1}{A}\sum\nolimits_{a\in\mathcal{A}}\mathrm{P}_a(y_a \mid \mathbf{y})\sum\nolimits_{v\in\overline{\mathcal{Y}}}\mathrm{P}(v, \mathbf{y}_{-a}) = \frac{1}{A}\sum\nolimits_{a\in\mathcal{A}}\mathrm{P}(\mathbf{y}) = \frac{A}{A}\mathrm{P}(\mathbf{y}) = \mathrm{P}(\mathbf{y}).$$

Also, the left hand side of (5) is

$$\mu(\mathbf{y}) = \mathrm{P}(\mathbf{y}).$$

Thus $\mu(\mathbf{y}) = \mathrm{P}(\mathbf{y})$ solves (5) for each $\mathbf{y} \in \overline{\mathcal{Y}}^A$.

As we mentioned earlier, and assumed in the last few lines, the existence of Gibbs equilibrium also requires the set of conditional probabilities $(\mathrm{P}_a)_{a\in\mathcal{A}}$ to be compatible. We formalize this idea next.

Definition: We say $(\mathrm{P}_a)_{a\in A}$ is a set of compatible conditional distributions if there exists a joint distribution $\mathrm{P} : \overline{\mathcal{Y}}^A \to [0,1]$, with $\sum_{\mathbf{y}\in\overline{\mathcal{Y}}^A} \mathrm{P}(\mathbf{y}) = 1$, such that

$$\mathrm{P}_a(y_a \mid \mathbf{y}) = \mathrm{P}(\mathbf{y}) / \sum\nolimits_{y_a\in\overline{\mathcal{Y}}} \mathrm{P}(\mathbf{y}) \text{ for each } \mathbf{y} \in \overline{\mathcal{Y}}^A.$$

The technical conditions required for a set of conditional distributions to be compatible are discussed in Kaiser and Cressie (2000). Their analysis implies that compatibility demands strong symmetric restrictions. In particular, in the two people, two actions case, Arnold and Press (1989) show that compatibility holds if and only if the next equality is satisfied

$$\frac{1-\mathrm{Q}_1(1\mid 0)}{\mathrm{Q}_1(1\mid 0)}\frac{\mathrm{Q}_1(1\mid 1)}{1-\mathrm{Q}_1(1\mid 1)} = \frac{1-\mathrm{Q}_2(1\mid 0)}{\mathrm{Q}_2(1\mid 0)}\frac{\mathrm{Q}_2(1\mid 1)}{1-\mathrm{Q}_2(1\mid 1)}.$$

The last result states that, under specific conditions, the Gibbs equilibrium coincides with μ. A similar connection is discussed in Blume and Durlauf (2003). The next proposition puts together all our previous ideas.

Proposition B.2. *Assume A1 and G1 hold. If $(\mathrm{P}_a)_{a\in A}$ is a set of compatible conditional distributions, then $\mathrm{P}_a(y_a \mid \mathbf{y}) = \mu(\mathbf{y}) / \mu_{-a}(\mathbf{y}_{-a})$ for each $\mathbf{y} \in \overline{\mathcal{Y}}^A$.*

C. Monte Carlo Simulations

This appendix describes how we generated the observations for the restaurant model.

Let $\lambda = \sum_{a\in\mathcal{A}} \lambda_a$. We generate the data according to an iterative procedure for a fixed time period $\mathcal{T}$. The k-th iteration of the procedure is as follows:

(i) Given $\mathbf{y}_{k-1}$ set $\mathbf{y}_k = \mathbf{y}_{k-1}$;

(ii) Generate a draw from the exponential distribution with mean $1/\lambda$ and call it x_k;

(iii) Randomly sample an agent from the set $\mathcal{A}$, such that the probability that a is picked is λ_a/λ;

(iv) Given the agent selected in the previous step and the current choice configuration $\mathbf{y}_k$ construct a consideration set using Q_a;

(v) If the consideration set is empty, then set $y_{a,k} = 0$. Otherwise pick the best alternative according to the preference order of agent a from the consideration set and assign it to $y_{a,k}$.

Given the initial configuration of choices $\mathbf{y}_0$ we applied the above algorithm till we reached $\sum_k x_k > \mathcal{T}$ (On average the length of the sequence is $\lambda\mathcal{T}$). Define $z_k = \sum_{l\leq k} x_l$. The continuous time data is $\{(y_k, z_k)\}$. The discrete time data is obtained from the continuous time data by splitting the interval $[0, \mathcal{T}]$ into $T = [\mathcal{T}/\Delta]$ intervals and recording the configuration of the network at every time period $t = i\Delta$, $i = 0, 1, \ldots, [\mathcal{T}/\Delta]$.