DITCH – A Model of Inter-Ethnic Partnership Formation

1. Purpose

The DITCH ("Diversity and Inter-ethnic marriage: Trust, Culture and Homophily") model has been developed to investigate the processes of partner selection in diverse communities, focusing on individual preferences, opportunities for contact, and group size to uncover how these may lead to differential rates of inter-ethnic marriage. Inter-ethnic marriage is both a cause and a consequence of immigrant integration. It is, however, unclear how individual preferences and opportunities for contact may combine to produce the spectrum of rates of inter-ethnic marriage we observe in the UK and elsewhere. The overall results from the DITCH model indicate that – as expected and in line with existing evidence – diversity (especially in areas with low ethnic homogeneity) fosters high rates of inter-ethnic marriage, but these can be mediated by group size, the type of social network and the extend of the partner search range.

2. Entities, state variables, and scales

Our approach in developing the model has been to start with a simple model that includes only processes and data essential to modelling (inter-ethnic) partnership formation, but which can be easily extended when necessary. The model is therefore constrained to simulating a cohort of agents aged 18-35 years who are single at model initialisation and search for a suitable partner within their social network.

In the current model version agents are characterised by five traits (implemented as a list with 5 elements):

1. Gender (male or female).
2. Age (18-35).
3. Ethnicity (one of four discrete values representing abstract ethnicities: w, x, y, z).
4. Compatibility (a real number between 0 and 1 modelling "chemistry" between agents; the closer their compatibility scores, the more compatible are the agents).
5. Education (an integer [0,4] representing education level with 0: none, 1: level 1 (some GCSEs), 2: level 2 (GCSEs), 3: level 3 (A-Levels), 4: level4/5 (university degrees)).

The first three traits are used to determine an agent's graphical representation: Their ethnicity is visualised as their colour (w: orange, x: green, y: yellow, z: cyan), their gender as their shape, and their age as their size (the older, the bigger).

Agents have preferences for partners with certain traits (another list with 5 elements). These are expressed as follows:

1. Gender: opposite gender. (We assume a strictly heterosexual world in this iteration of the model for the sake of simplicity.)
2. Age: a range with a mean of 1.3 years and a standard deviation of 6.34 years (see, e.g., Hirsch et al 2010) centred on the agent's own age.

3. Ethnicity: a preference value for each of the four represented ethnicities, with a slight bias towards the agent's own ethnicity. These are initialised randomly with values between 0 and 1; making sure that the agent's own ethnicity receives the highest value .

4. Compatibility: a range of +- 0.05 to 0.5 centred on the agent's own compatibility value. This is initialized using as a random uniform distribution.

5. Education: Instead of having a preference for a particular education level, we assume that agents prefer their potential partner to not differ too much from their own education level . Education preference is therefore expressed as a tolerance for the difference in education level (from 0 to 4). This is initialised via a normal distribution with mean 0 and standard deviation between 0 and 4 (set via the model parameter *sd-education-pref*).

Agents also have social links to other agents. Since agents are 18-35 years old at model initialisation, at least some of their social links need already be established. These can be ties with friends, family, neighbours or colleagues; the model does not differentiate between different types of links. The process of establishing the initial social network is described in section 5. Initialisation.

Further agent characteristics exist to simplify programming, collect statistics or aid debugging. These are not described here.

The agents live in a purely social environment. The 2D grid world, in which they are positioned, does not represent the "real" world and is only used to form the social network. After the initialisation phase, their position on the grid does not change during the simulation.

3. Process overview and scheduling

Time is modelled in discrete steps ("ticks"). The length of each step can be controlled via the model parameter *ticks-per-year*. Default is 12, i.e. a monthly time step. After the model initialisation, the following processes are executed each time step, in this order:

1. *Partner search*: All single agents, i.e. agents, who are not currently dating someone or are already married, search for a potential partner within their social network.

   This means they filter their (unmarried / not dating) friends for the preferred gender and age range. In this model version, agents apply the "dynamic love-nipple-based approach", i.e. they start looking in their direct links (level 1) before extending the search outwards (level 2+) with the model parameter *love-radar* dictating the range of the search. Each tick they will pick up to 3 potential partners to choose randomly from. This search approach is derived from the 'area-restricted search' literature as outlined in (Hills and Todd 2008).

   If another agent has already tagged an agent as a potential partner, this agent will not look for a different potential partner during this tick. This ensures that each agent will only have up to one blind date per tick.

   To avoid finding and dating the same potential partner over and over again, agents keep track of everyone they recently dated and are only allowed to pick one of them again after a specified period of time (model parameter *second-chance-interval*).

2. *Blind date*: All single agents, who found a potential partner in the previous stage, go on a date with this partner to determine, if they are satisfied with each other. An agent is satisfied with their potential partner if their "visible" traits fall within the agent's preference ranges, i.e.:

   a. The partner is of the preferred gender and within the preferred age range (already established during partner search).

   b. The partner's ethnicity is "acceptable". This is determined by interpreting the preference values for ethnicities as probabilities. If a random number between 0,1 is smaller than the preference value, then the partner's ethnicity is acceptable.

   c. The difference in partner's education level and own education level is <= the preferred difference.

3. *Start dating*: If both agents on a blind date are satisfied with their potential partner, they start dating, i.e. they will not continue to look for potential partners. The dating period is determined by a normal distribution whose shape can be set via the model parameters *mean-dating* and *sd-dating*.

4. Proposal: Agents, who have reached the end of their dating period, check if they want to marry each other. This involves the "invisible" trait compatibility: If the two agents are compatible, i.e. their compatibility scores both lie within the other's preferred range, they will marry with a probability of 1 – |compatibilitya1 – compatibilitya2|. Thus agents have a higher chance of partnering the closer their compability scores. This process is derived from 'the mate searching game' (Todd 1997; Gilbert and Troitzsch 2005, 2–3). If agents partner successfully, their link becomes visible. They remain in the model to keep the social network intact, but are not available as potential partners for others anymore.

5. *Preference update*: All single agents with unsuccesful or no dates update their preferences. They expand their preferred age range by 0.1 and reduce their preference value for the partner's ethnicity by 0.01. Preferences are only updated after a number of unsuccessful dates. This number is specified as model parameter *update-threshold*.

   Incompatible agents, i.e. dating agents, who tried unsuccesfully for marriage, extend their preferred compatibility range by 0.05 if the update threshold is reached.

6. *Grow network*: Agents' social networks are dynamic, i.e. they change during the simulation. This is represented in an abstract way since the model in its current version focusses on the process of partnership formation and does not contain any other social processes that would influence the formation of new links like meeting at work, leisure activities or places of worship.

   Each agent has the chance to form a new link with either (95%) the most similar friend of an existing friend or (5%) a randomly picked stranger. Similarity is determined with regard to age, ethnicity and education; gender is not taken into account. The probability of link formation is controlled via the model parameter *new-link-chance* and was set to 0.5 (50%) in our experiments. With a small probability of 5%, each agent may also drop a random social link (except for the link to their spouse).

At the end of each time step, output variables are updated and each year (ticks modulo *ticks-per-year* == 0) agents age. For validation purposes the model will also produce a snapshot of the social network every 5 years, saved as lists of nodes and edges in two comma-separated files, if the global variable *file-output?* is set to true.

4. Design concepts

Basic principles. Partner selection, and its determinants, has a long been a topic of enquiry in social science. In our model we draw the rules governing the agents' behaviour on existing sociological evidence stating that both individual-level preferences (via assortative mating (Becker 1973)), opportunities for contact (via diversity, group size, population size and sex ratios (Blau et al. 1982, Chiswick 2009)), and family and kin networks (Ballard 2008, Gardner 2006) are important drivers of inter-ethnic partner choice.

Emergence. The key model outputs – the number and composition of (interethnic) marriages – is an emergent feature of the agents' behaviour.

Adaptation. While the agents' preferences for partners are somewhat adapted in response to their experiences (see section *Learning* below), the agents' behaviour itself does not adapt.

Objectives. The agents' sole objective in this model is to find a suitable partner, i.e. a partner whose traits lie within the agent's preferences. Agents are satisficing in that they do not aim for the "best" or "perfect" match but will settle with a partner within their range of preferences.

Learning. Agents learn from their dating experience and adapt their preferences accordingly. Unsuccessful dates or marriage proposals lead agents to lower their expectations in that they widen their preferred age range (by 0.1) or their compatibility range (by 0.05), respectively. They also have a negative impact on the agents' ethnicity preferences in that the preference for their "rejected" potential partner's ethnicity is reduced by 0.01. Successful dates, on the other hand, boost the preference for the potential partner's ethnicity by 0.01.

Prediction. Agents do not use prediction in this model.

Sensing. This is a purely social model thus agents do not sense their environment.

Interaction. Agents interact during their "blind" dates, when they are exchanging their "visible" characteristics to determine if they are satisfied with their potential partner, and during the proposal phase, when they are exchanging their "invisible" characteristics (compatibility score) to determine if they are compatible for a partnership.

Stochasticity. For lack of available data, a number of processes are modelled using stochastic elements. These are foremost the formation of the social network (picking neighbours to form initial links with, picking a random friend to form links with the most similar friend of this friend, dropping a random link) and the choice of potential partner (picking a random agent from a pre-selected set of up to 3 agents). There is also a random element in the proposal process to avoid having all agents settle with the first compatible partner they encounter.

Further stochasticity is used during the initialisation of the model. The agents' traits age, ethnicity and education follow empirically defined stochastic distributions, whereas gender and compatibility are assigned using random uniform and random normal distributions, respectively.

Collectives. There are no collectives in this model.

Observation. The model uses four global variables to compute the following output measures:

- num-turtles: array of the number of turtles per ethnicity (this is static in this model version)
- married-turtles: array of the number of married turtles per ethnicity
- out-percent: array of the percent of cross-ethnic marriages per ethnicity

- cross-ethnic: percent of cross-ethnic marriages overall

These measures have been used in batches of simulation experiments run via NetLogo's BehaviourSpace.

The model also provides graphs and monitors to be observed during a simulation run. These can be divided into observations related to the evolution of the social network and observations related to the formation of partnerships. The former are:

- Link Distribution: Histogram of the number of social links agents currently have
- Proportion of inter-ethnic links: Time series per ethnicity (w: orange, x: green, y: yellow, z: cyan) of the proportion of inter-ethnic links in the social networks of agents
- count links: Total number of links in the social network
- Number of Links: Time series of the total number of links (black) and the total number of dropped links (grey) in the social network

The latter are:

- marriages: Total number of marriages
- cross-ethnic %: Total proportion of inter-ethnic marriages
- max cross-ethnic %: The maximal possible proportion of inter-ethnic marriages given the ethnic composition of the population
- # dates before marriage: Average number of blind dates an agent has before marrying a partner
- Age at Marriage: Histogram of the age agents are when they marry
- % married: Time series of the proportion of married agents per ethnicity
- Age of Singles: Histogram of the age of single agents
- % single: Time series of the proportion of single agents per ethnicity
- Dates (Distance 1, Distance 2, Distance 3): Total number of dates with partners of a social distance of 1, 2, or 3, respectively.
- Marriages (Distance 1, Distance 2, Distance 3): Total number of marriages with partners of a social distance of 1, 2, or 3, respectively. This and Dates is to show the effect of the parameter *love-radar*.
- Overall marriage % (w-w, w-x, w-y, w-z, x-x, x-y, x-z, y-y, y-z, z-z): Overall proportion of marriages between partners of the named ethnicities; e.g.: x-y shows the number of marriages between agents of ethnicity x and agents of ethnicity y divided by the total number of marriages.
- Group % (w-in, w-out, x-in, x-out, y-in, y-out, z-in, z-out): Proportion of intra- and inter-ethnic marriages per ethnicity, e.g. w-in shows the proportion of marriages within ethnicity w (number of marriages between agents of ethnicity w divided by number of marriages with at least one partner of ethnicity w), whereas w-out shows the proportion of inter-ethnic marriages for ethnicity w (number of marriages between an agent of ethnicity w and an agent of a different ethnicity divided by number of marriages with at least one partner of ethnicity w).

5. Initialisation

Agent traits are initialised as follows:

1. *Gender:* This is initialised randomly with 50% chance each of being male or female.

2. *Age*: This is initialised according to an age distribution taken from the UK Census 2001:

Age	18-19	20-24	25-29	30-34	35-36
Proportion	9.99%	24.98%	27.48%	31.29%	6.26%

Within each interval, age is assigned uniformly randomly.

3. *Ethnicity*: This is initialised randomly according to the proportions specified via the model parameter *eth-proportions*. Each abstract ethnicity (w, x, y, z) is assigned a particular ethnic group via the model parameters *ethnicity-w*, *ethnicity-x*, *ethnicity-y*, *ethnicity-z*. The following ethnic groups are available: White British, White Irish, Other White, White and Black Caribbean, White and Black African, White and Asian, Other Mixed, Indian, Pakistani, Bangladeshi, Other Asian, Black Caribbean, Black African, Other Black, Chinese, Other Ethnic Group.

4. *Compatibility:* This is initialised as a normal distribution with mean 0.5 and standard deviation 0.2 to achieve normally distributed values between 0 and 1.

5. *Education:* This is initialised according to distributions particular for ethnic group and gender taken from the 2001 Census in the age range of our model.

The initial values for agent preferences have been given in section 2.

The social network is initialised as follows: At creation, agents are trying to find an empty cell on the grid next to one of their own ethnicity. Only if that is not possible, they pick a random free cell. This results in clustering of minorities. Agents then form links with some of their neighbours. After this, agents segregate according to age for 50 ticks and then form links with some of their neighbours again. This attempts to achieve a social network where agents have cross-ethnic links to agents of a similar age ("school friends") and same-ethnic links to agents of a wider age range ("family", "neighbours").

In the very first model version, agents formed links with all of their neighbours. Since the resulting network seemed too dense, agents now only form links with 40-50% of their neighbours with a slight preference (= higher chance) for same gender. In detail, for each neighbour there is a chance of 40% (50% if same gender) to form a link.

The number and ethnic composition of agents depends on the model parameters *num-agents*, *eth-proportions* and *ethnicity-w*, *ethnicity-x*, *ethnicity-y*, *ethnicity-z.* We have run the model with 1000 agents for different scenarios derived from sample areas representing different typologies of ethnic diversity within local and urban authorities in England & Wales in 2001, based on the number of ethnic groups and the proportion of the population from a 'White: British' background. Within the sample areas the four largest ethnic groups were selected (if over 1% of the overall population) and their proportions weighted to add up to 100%. The sample areas are as follows (with weighted ethnic group proportions):

1. *Super-diverse*: Newham, London

   Sample population: White: British (WB): 49.8%; Asian/Asian British: Indian (A/ABI): 17.9%; Asian/Asian British: Bangladeshi (A/ABB): 13.0%; Black/Black British: African (B/BBA): 19.3%

2. *Cosmopolitan*: Trafford, Greater Manchester

   Sample population: WB: 93.2%; White: Irish: 2.8%; A/ABI: 2.1%; Asian/Asian British: Pakistani (A/ABP): 1.5%

3. *Bifurcated*: Bradford, W. Yorkshire

   Sample population: WB: 80.3%; White: Other (WO): 1.6%; A/ABI: 2.8%; A/ABP: 15.3%

4. *Parochial*: Cheshire West & Chester

   Sample population: WB: 98.2%; WO: 1.8%

The other model parameters have been set to the following values for these experiments:

Parameter name	*Parameter description*	*Value*
sd-education-pref	Standard deviation of the normal distribution governing the agents' preference for difference in education level (mean is always 0).	1
update-threshold	Number of (unsuccessful) dates necessary before an agent changes its preference values.	5
mean-dating	Mean and standard deviation (in years) of the normal distribution governing the duration of the agents' dating period.	1
sd-dating		1
ticks-per-year	Number of ticks (time steps) per year. The chosen value of 12 results in a monthly time step.	12
new-link-chance	The probability of an agent forming a new link per tick.	0.5
second-chance-interval	The number of ticks an agent will wait before considering a previously dated agent as a potential partner again.	6

6. Input data

The model does not use input data to represent time-varying processes.

7. Submodels

There are no submodels per se in this model but the model processes as described in section 3 can be divided into (a) partnership formation (partner search, blind date, start dating, proposal and preference update) and (b) social network formation (grow network and its initialisation, see section 5).

References

R. Ballard (2008): 'Inside and outside: contrasting perspectives on the dynamics of kinship and marriage in contemporary South Asian transnational networks'. In R. Grillo (ed.), The Family in Question: Immigrant Ethnic Minorities in Multicultural Europe. Amsterdam: Amsterdam University Press, pp. 37-70.

G. S. Becker (1973): 'A theory of marriage: Part I'. Journal of Political Economy, 81(4), pp. 813–846.

P.M. Blau, T.C. Blum, and J.E. Schwartz (1982): 'Heterogeneity and inter-marriage'. American Sociological Review, 47(1), pp. 45–62.

C.U. Chiswick (2009): 'The economic determinants of ethnic assimilation'. Journal of Population Economics, 22(4), pp. 859–880.

K. Gardner (2006): 'The transnational work of kinship and caring: British Bangladeshi marriages in historical perspective'. Global Networks, 6(4), pp. 373-389.

N. Gilbert and K. Troitzsch (2005): 'Simulation for the Social Scientist', 2nd ed. Maidenhead: Open University Press.

T. Hills and P.M. Todd (2008): 'Population heterogeneity and individual differences in an assortative agent-based marriage and divorce model (MADAM) using search with relaxing expectations'. Journal of Artificial Societies and Social Simulation 11(4)5. <http://jasss.soc.aurrey.ac.uk/11/4/5.html>

P. M. Todd (1997): 'Searching for the next best mate'. In R. Conte, R. Hegselmann and P. Terna (eds.), Simulating Social Phenomena. Berlin: Springer, pp. 419-436.