Computing the Complexity for Schelling Segregation Models

Stefan Gerhold
RISC, J. Kepler University
Altenberger Str. 69
A–4040 Linz, Austria

Lev Glebsky
IICO-UASLP
Av. Karakorum 1470
Lomas 4a.
78210 San Luis Potosi
SLP México

Carsten Schneider *
RISC, J. Kepler University
Altenberger Str. 69
A–4040 Linz, Austria

Howard Weiss †
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30334
USA

Burkhard Zimmermann‡
RISC, J. Kepler University
Altenberger Str. 69
A–4040 Linz, Austria

April 26, 2007

Abstract

The Schelling segregation models are "agent based" population models, where individual members of the population (agents) interact directly with other agents and move in space and time. In this note we study one-dimensional Schelling population models as finite dynamical systems. We define a natural notion of entropy which measures the complexity of the family of these dynamical systems. The entropy counts the asymptotic growth rate of the number of limit states. We find formulas and deduce precise asymptotics for the number of limit states, which enable us to explicitly compute the entropy.

*Supported by the SFB-grant F1305 and the grant P16613-N12 of the Austrian FWF.
†This work was partially supported by NSF grants DMS-0355180 and INT-0104675
‡Supported by the SFB-grant F1301 of the Austrian FWF.

1 Introduction

In this note we study the dynamics of one-dimensional Schelling population models. The Schelling models [1, 2, 3] are "agent based" population models, where individual members of the population (agents) interact directly with other agents and move in space and time. In the vast majority of population models, e.g., all ordinary differential equation and partial differential equation models, sub-populations interact with other sub-populations with strong mixing assumptions, and the individuals are only implicit in the model. Agent based models such as the Schelling model are fundamentally different, and are becoming an important tool in population and disease modeling.

The Schelling model serves as a paradigm for modeling population movement via non-local aggregation. Devised nearly thirty years ago, this model is still actively studied and heavily cited by demographers trying to understand the relationship between residential choices of individuals and aggregate patterns of neighborhood and city change. Schelling received the 2006 Nobel Prize in Economics and this model was cited by the Nobel committee. Besides the manuscript [6], we are not aware of any rigorous results on this general family of models. In fact, there appear to be very few rigorous results in the entire area of agent-based modeling.

In this manuscript we study the one-dimensional Schelling segregation model, both on a line of L lattice sites and on a circle of L lattice sites, as finite dynamical systems. Measures of "orbit complexity" of a dynamical system play a fundamental role in nonlinear dynamics. Accordingly, we define a natural notion of entropy which measures the complexity of the family of these dynamical systems. The entropy counts the asymptotic growth rate of the number of limit states. We find formulas and deduce precise asymptotics for the number of limit states, which enable us to explicitly compute the entropy.

More precisely, for each L, we obtain explicit formulas for the number of limit states. These formulas involve sums of hypergeometric series, and appear to have no closed form expressions. In Section 6 we compute the precise asymptotics for the model on a line. To do so, we calculate the generating function of the dominant part of the number of limit states and analyse it with a complex analysis based asymptotic method.

An analogous enumeration of the limit states for the two-dimensional Schelling models on a $L \times L$ lattice seems extremely difficult, and may be a NP hard problem.

2 Description of the one-dimension model

We consider natural one-dimensional versions of the Schelling model on an interval and on a circle (i.e., with periodic boundary conditions). In the first case the configuration space is $\Lambda_L = \{1, 2, \cdots, L\}$ and in the second case the configuration space is $\mathbb{Z}_L = \mathbb{Z}/L\mathbb{Z}$. In both cases each lattice site is inhabited by an agent having one of two distinct labels, which we denote by A and B. We assume that M sites are filled with $A's$ and N sites are filled with $B's$, where $L = M + N$. This describes a *configuration* or *state* of the system. For an example of the model on $\mathbb{Z}_{M+N}$, one can consider M Democrats and N Republicans seated around a poker table.

Given a state, we say that a site is *happy* provided the label at that site is not isolated, i.e., at least one of its two nearest neighbors has the same label. For the model on an interval, an agent on the boundary is happy provided the one nearest neighbor has the same label.

To describe the evolution of a state x_n, one selects two sites having different labels. This can be done following a deterministic algorithm or randomly. If *both* chosen agents are unhappy, then we interchange the two labels, and this defines state x_{n+1}. The interchange makes both chosen agents happy. If at least one of the chosen agents is already happy, we select again, until we have selected two unhappy agents. We continue this procedure until we reach a state for which it is impossible to choose two isolated agents. We call such a state a *limit state*. Although different choices will lead to different limit sets, the number of limit sets is independent of these selection choices. Since at each step the chosen sites can be far apart and only two randomly chosen sites can update at each step, this is quite different than the usual nearest neighbor models in statistical physics or cellular automata (CA) models.

Some limit sets have no isolated agents while others have one or more isolated agents with the same label. In the first case, limits sets correspond to alternating groups of agents, where each group consists of at least two agents with the same label. The latter case is similar, except that, in addition, there are some isolated agents all with the same label. Thus some agents may be unhappy, but they have nowhere to move

3 Brief remarks on the dynamics of the system

The Schelling model can be viewed as a dynamical system on configurations on a finite lattice. With stochastic selection rules, we view the Scehlling model as a random dynamical system. There exists a global Lyapunov function L on the configurations, i.e., a function defined on configurations that strictly decreases in time until a limit state is reached. For the circular lattice, each possible global configuration of A's and B's is specified by a function $x\colon \mathbb{Z}_{M+N} \to \{-1, 1\}$, where one associates label 1 for agent A and -1 for agent B. It is easy to show that the function

$$L(x) = -\sum_{v \in \Lambda_N} \sum_{i=\{-1,1\}} x(v) \cdot x(v+i)$$

is strictly decreasing along the evolution of the system until it reaches a limit state. See [6] for details, along with four interpretations of the function L. A similar Lyapunov function can be constructed for the dynamics on Λ_{M+N}. The number of sites labeled A (or B) does not change during the evolution of the system, so in this sense this dynamical system has a conserved quantity.

There is clearly no recurrence in these dynamical systems. In fact, for random selection rules this system can be described by a transient Markov chain. A natural way to quantify the complexity of this family of dynamical systems is through the asymptotic growth rate of the number of limit states as a function of lattice size. Theorem 5 and Proposition 3 contain explicit formulas for the number of limit states, and Theorem 5 provides precise asymptotics for the number of limit states on the line. This lets us explicitly compute the entropy.

4 Schelling model on $\{1, 2, \cdots, M+N\}$

We assume that the configuration space is $\Lambda_{M+N} = \{1, 2, \ldots, M+N\}$ and the sequence of M $A's$ and N $B's$ begins with a group of $A's$, ends with a group of $B's$, and consists of m_1 $A's$ followed by n_1 $B's$ followed by m_2 $A's$ followed by n_2 $B's$, ..., followed by m_k $A's$ followed by n_k $B's$, where each m_i is at least one and each n_i is at least two. The following are two examples:

$$\underbrace{AAAA}_{m_1}\underbrace{BB}_{n_1}\underbrace{AAA}_{m_2}\underbrace{BBBB}_{m_2}\cdots\underbrace{AAAA}_{m_k}\underbrace{BBB}_{n_k}$$
$$\underbrace{A}_{m_1}\underbrace{BBB}_{n_1}\underbrace{AA}_{m_2}\underbrace{BB}_{m_2}\cdots\underbrace{A}_{m_k}\underbrace{BB}_{n_k}.$$

Each such limit state corresponds to a multiset $[m_1 n_1 m_2 n_2 \ldots m_l n_k]$ satisfying the following conditions:

$$\left\{\begin{array}{l} m_1 + m_2 + \cdots + m_k = M \\ n_1 + n_2 + \cdots + n_k = N \\ m_i \geq 1 \quad i = 1, 2, \ldots, k \\ n_i \geq 2 \quad i = 1, 2, \ldots, k \end{array}\right\}. \tag{1}$$

We denote $\mathcal{M}(L, k, r)$ to be the collection of multisets of the form $[m_1 m_2 \ldots m_k]$ where $m_1 + m_2 + \cdots + m_k = M$ and $m_i \geq r$, $i = 1, 2, \ldots, k$. An elementary counting argument shows that

$$\#\mathcal{M}(M, k, r) = \binom{M + (1-r)k - 1}{k-1},$$

and thus the total number of multisets of the form $[m_1 n_1 m_2 n_2 \ldots m_l n_k]$ which satisfy (1) is

$$\#\mathcal{M}(M, k, 1)\,\#\mathcal{M}(N, k, 2) = \binom{M-1}{k-1}\binom{N-k-1}{k-1}. \tag{2}$$

Since the total number of these limit sets corresponds to the sum over k of the number of multisets of the form $[m_1 n_1 m_2 n_2 \ldots m_l n_k]$ which satisfy (1), we obtain that the number of limit sets for the Schelling model on $\{1, 2, \cdots, M+N\}$

which begin with A, end with B, and can have isolated $A's$ is

$$\sum_{k=1}^{M+N} \binom{M-1}{k-1} \binom{N-k-1}{k-1}.$$

The same formula holds for the number of limit states which begin with B, end with A, and which can have isolated $A's$. *These sums do not seem to have closed form expressions.*

We now count those limit sets which begin and end with A and which can have isolated $A's$. From (2) it follows that the number of multisets $[m_1 n_1 m_2 n_2 \ldots m_l n_k m_{k+1}]$ satisfying:

$$\left\{ \begin{array}{l} m_1 + m_2 + \cdots + m_{k+1} = M \\ n_1 + n_2 + \cdots + n_k = N \\ m_i \geq 1 \quad i = 1, 2, \ldots, k+1 \\ n_i \geq 2 \quad i = 1, 2, \ldots, k. \end{array} \right\} \tag{3}$$

is

$$\binom{M-1}{k} \binom{N-k-1}{k-1}.$$

To compute the total number of limit states one must count the following twelve types of limit states:

Name	Starting agent	Ending agent	May have isolated	Total number
TA_1	A	B	A	$\sum \binom{M-1}{k-1}\binom{N-K-1}{k-1}$
TB_1	A	B	B	$\sum \binom{M-k-1}{k-1}\binom{N-1}{k-1}$
TA_2	B	A	A	$\sum \binom{M-1}{k-1}\binom{N-K-1}{k-1}$
TB_2	B	A	B	$\sum \binom{M-k-1}{k-1}\binom{N-1}{k-1}$
TA_3	A	A	A	$\sum \binom{M-1}{k-1}\binom{N-k-1}{k-1}$
TB_3	A	A	B	$\sum \binom{M-1}{k-1}\binom{N-k-1}{k-1}$
TA_4	B	B	A	$\sum \binom{M-1}{k-1}\binom{N-k-1}{k-1}$
TB_4	B	B	B	$\sum \binom{M-1}{k-1}\binom{N-k-1}{k-1}$
T_1	A	B	No	$\sum \binom{M-1}{k-1}\binom{N-k-1}{k-1}$
T_2	B	A	No	$\sum \binom{M-1}{k-1}\binom{N-k-1}{k-1}$
T_3	A	A	No	$\sum \binom{M-1}{k-1}\binom{N-k-1}{k-1}$
T_4	B	B	No	$\sum \binom{M-1}{k-1}\binom{N-k-1}{k-1}$

The total number of limit states on $\{1, 2, \cdots, M+N\}$ with M number of $A's$ and N number of $B's$ is $|\cup_i (TA_i \cup TB_i)|$. Since the only nonempty intersections are $TA_i \cap TB_i = T_i$, we have $|\cup_i (TA_i \cup TB_i)| = \sum_i (|TA_i| + |TB_i| - |T_i|)$. We thus obtain the following theorem.

Theorem 1. *The total number of limit states for the Schelling segregation model on* $\{1, 2, \cdots, M+N\}$ *with* M *number of* $A's$ *and* N *number of* $B's$ *is*

$$\begin{aligned}
s(M,N) = {} & 2\sum_{k=1}^{M+N}\binom{M-1}{k-1}\binom{N-k-1}{k-1} + 2\sum_{k=1}^{M+N}\binom{M-k-1}{k-1}\binom{N-1}{k-1} \\
& + \sum_{k=1}^{M+N}\binom{M-1}{k}\binom{N-k-1}{k-1} + \sum_{k=1}^{M+N}\binom{M-k-2}{k}\binom{N-1}{k-1} \\
& + \sum_{k=1}^{M+N}\binom{M-1}{k-1}\binom{N-k-2}{k} + \sum_{k=1}^{M+N}\binom{M-k-1}{k-1}\binom{N-1}{k} \\
& -2\sum_{k=1}^{M+N}\binom{M-k-1}{k-1}\binom{N-k-1}{k-1} - \sum_{k=1}^{M+N}\binom{M-k-2}{k}\binom{N-k-1}{k-1} \\
& - \sum_{k=1}^{M+N}\binom{M-k-1}{k-1}\binom{N-k-2}{k}.
\end{aligned}$$

Corollary 2. *When there are equal numbers of agents with both labels, i.e.,* $M = N$*, the formula in Theorem 1 reduces to*

$$\begin{aligned}
s(N,N) = {} & 4\sum_{k=1}^{2N}\binom{N-1}{k-1}\binom{N-k-1}{k-1} + 2\sum_{k=1}^{2N}\binom{N-1}{k}\binom{N-k-1}{k-1} \\
& +2\sum_{k=1}^{2N}\binom{M-k-2}{k}\binom{N-1}{k-1} - 2\sum_{k=1}^{2N}\binom{N-k-1}{k-1}\binom{N-k-1}{k-1} \\
& -2\sum_{k=1}^{2N}\binom{N-k-2}{k}\binom{N-k-1}{k-1}.
\end{aligned} \tag{4}$$

Each sum in the above expression for $s(M,N)$ can be written as a hypergeometric series. In Section 6 we study the asymptotics of $s(N,N)$. We show that the dominant part of $s(N,N)$ consists of the first three sums in (4), derive a recurrence for the dominant part, find its generating function, and deduce that $s(N,N) \sim \text{const} \cdot 3^N/\sqrt{N}$.

5 Schelling model on the circular lattice $\mathbb{Z}_N$

Suppose that all the Democrats and Republicans sitting around a round poker table each advance k seats in a fixed direction. After this seating rearrangement, nobody's neighbors have changed, and thus we do not wish to consider the two configurations as distinct. Hence to count the number of limit states for the Schelling model on the circular lattice $\mathbb{Z}_N$ we need to factor out rotational symmetries.

The cyclic group $\mathbb{Z}_k$ acts on the following set of pairs of integers:

$$\{((n_1, m_1), (n_2, m_2), ..., (n_k, m_k)) \mid n_i, m_i \geq 2, \ \sum_{i=1}^{k} n_i = N, \ \sum_{i=1}^{k} m_i = M\}$$

by cyclic shift of the indexes. We need to calculate the number of orbits of this action and then sum it over k. Pólya's method to count the number of orbits of a finite group G acting on a set can be expressed by the following formula:

$$\#\text{Orbits} = \frac{1}{|G|} \sum_{g \in G} \#\text{Fix}(g), \tag{5}$$

where $\text{Fix}(g)$ denotes the fixed point set for g.

For our case $r \in \mathbb{Z}_k = \{1, 2, ..., k\}$ (to avoid $\gcd(0, k)$, we take $r = k$ instead of $r = 0$). Let $\gcd(r, k) = c$ and $d = k/c$. Then

$$\#\text{Fix}(r) = \begin{cases} 0 & \text{if } d \nmid N \text{ or } d \nmid M \\ \binom{\frac{N-k}{d}-1}{\frac{k}{d}-1}\binom{\frac{M-k}{d}-1}{\frac{k}{d}-1} & \text{if } d|N \text{ and } d|M. \end{cases}$$

It is convenient to introduce the Euler totient function $\phi(n) = \#\{1 \leq m \leq n \mid \gcd(n, m) = 1\}$. It is easy to check that $\phi(k/c) = \#\{1 \leq r \leq k \mid \gcd(r, k) = c\}$. Using this fact and (5) we obtain:

$$\#\text{Orbits} = \frac{1}{k} \sum_{d| \gcd(k,M,N)} \phi(d) \binom{\frac{N-k}{d} - 1}{\frac{k}{d} - 1} \binom{\frac{M-k}{d} - 1}{\frac{k}{d} - 1}.$$

We sum this expression over k from 1 to $\min(M/2, N/2)$ and obtain

$$B(M, N) = \sum_k \sum_{d| \gcd(k,M,N)} \frac{1}{k} \phi(d) \binom{\frac{N-k}{d} - 1}{\frac{k}{d} - 1} \binom{\frac{M-k}{d} - 1}{\frac{k}{d} - 1}.$$

Substituting $k = rd$, we can rewrite this expression as

$$\begin{aligned} B(M, N) &= \sum_r \sum_{d| \gcd(M,N)} \frac{\phi(d)}{rd} \binom{\frac{N}{d} - r - 1}{r - 1} \binom{\frac{M}{d} - r - 1}{r - 1} \\ &= \sum_{d| \gcd(M,N)} \frac{\phi(d)}{d} \sum_r \frac{1}{r} \binom{\frac{N}{d} - r - 1}{r - 1} \binom{\frac{M}{d} - r - 1}{r - 1} \\ &= \sum_{d| \gcd(M,N)} \frac{\phi(d)}{d} A\left(\frac{M}{d}, \frac{N}{d}\right), \end{aligned}$$

where

$$A(M, N) = \sum_k \frac{1}{k} \binom{N - k - 1}{k - 1} \binom{M - k - 1}{k - 1}.$$

Summarizing, we obtain the following result:

Proposition 3. *The total number of limit states for the Schelling model on* $\mathbb{Z}_{M+N}$ *with* M *number of* $A's$ *and* N *number of* $B's$ *is*

$$\sum_{d|\,\gcd(M,N)} \frac{\phi(d)}{d} \sum_r \frac{1}{r} \binom{\frac{N}{d} - r - 1}{r-1} \binom{\frac{M}{d} - r - 1}{r-1}.$$

Remarks

(i) $B(M,N)$ and $A(M,N)$ seem to have the same asymptotics, but $A(M,N)$ is not always an integer.

(ii) The totient function ϕ appears in combinatorics, see problem 27, chapter 1, Stanley "Enumerative combinatorics", v1.

Corollary 4. *If* $(M,N) = 1$, *then the total number of limit states for the Schelling model on* $\mathbb{Z}_N$ *is*

$$2N \sum_{k=1}^{N} \binom{M}{k-1} \binom{N-k-1}{k-1}.$$

6 The asymptotics of $s(N,N)$

In this section we obtain the precise asymptotic expression for $s(N,N)$, the number of limit states of the model on $\{1,2,\ldots,2N\}$ with an equal number $M = N$ of $A's$ and $B's$. By Corollary 2, $s(N,N)$ can be expressed by the sums

$$s_1(N) := \sum_{k\geq 1} \binom{N-1}{k-1} \binom{N-k-1}{k-1},$$

$$s_2(N) := \sum_{k\geq 1} \binom{N-1}{k} \binom{N-k-1}{k-1},$$

$$s_3(N) := \sum_{k\geq 1} \binom{N-1}{k-1} \binom{N-k-2}{k},$$

$$s_4(N) := \sum_{k\geq 1} \binom{N-k-1}{k-1}^2,$$

$$s_5(N) := \sum_{k\geq 1} \binom{N-k-2}{k} \binom{N-k-1}{k-1},$$

where binomial coefficients with negative upper entry are defined to be zero.

Theorem 5. *The sequence is*

$$s(N,N) = 4s_1(N) + 2\left(s_2(N) + s_3(N) - s_4(N) - s_5(N)\right)$$

satisfies

$$s(N,N) = 3^N \left(CN^{-1/2} + \mathrm{O}(N^{-1})\right), \quad \text{where} \quad C = \tfrac{3}{4}\sqrt{\tfrac{3}{\pi}} \approx 0.7329. \tag{6}$$

For the Schelling system on a line with equal numbers N of agents of both labels, a natural notion of entropy measures the exponential growth rate of the total number of limit states, i.e,

$$h = \lim_{N\to\infty} \frac{1}{N} \log s(N,N).$$

Theorem 5 allows us to explicitly compute this entropy for the Schelling system.

Corollary 6. *The entropy* $h = \log 3$.

In order to prove the theorem, we are going to analyze the generating functions of $s_i(N)$. The proof uses methods from symbolic summation and asymptotic results. By the Cauchy-Schwarz inequality and the well known identities

$$\sum_{k=0}^{N} \binom{N}{k} = 2^N \qquad \text{and} \qquad \sum_{k=0}^{N} \binom{N-k}{k} = \text{Fibonacci}(N+1),$$

the $s_i(N)$ are all of at most exponential growth, and therefore their generating functions $S_i(z) := \sum_{N\geq 0} s_i(N) z^N$ are analytic at $z = 0$.

The functions $S_i(z)$ are obviously not entire. The growth of their coefficients is intimately connected with the behaviour of the functions at their singularities, to be elaborated in what follows. If $z = z_i$ denotes the singularity of $S_i(z)$ that is closest to the origin, then it is a well-known fact from complex analysis that the coefficient of z^N in $S_i(z)$ satisfies $s_i(N) = \mathrm{O}((1/|z_0| + \varepsilon)^N)$, with ε an arbitrary positive real.

We now estimate $|z_i|$ for $i = 4, 5$. Let us begin by computing[1] the recurrence

$$\begin{aligned}(N+2)s_4(N+4) - (2N+3)s_4(N+3) - (N+1)s_4(N+2)& \\ - (2N+1)s_4(N+1) + Ns_4(N) = 0,& \quad N \geq 3.\end{aligned}$$

Here various summation packages are on the market that can do the job. e.g., one can take the Mathematica package `Zb` [8], an efficient implementation of Zeilberger's algorithm for hypergeometric summation [9]; in this article we used

[1]Note that we not only can compute the recurrence, but also obtain a proof for its correctness; these aspects will be considered in details for a similar but more sophisticated example in Proposition 7.

the summation package `Sigma` [7]. From the recurrence one can produce an LODE (Linear Ordinary Differential Equation)

$$\begin{aligned} &24(-1+5z)S_4'(z)+12(-1-13z+20z^2)S_4''(z)\\ &+4(-3-7z-33z^2+30z^3)S_4^{(3)}(z)+(2-13z-11z^2-31z^3+20z^4)S_4^{(4)}(z)\\ &\qquad\qquad\qquad\qquad+(z-2z^2-z^3-2z^4+z^5)S_4^{(5)}(z)=0 \end{aligned}$$

for the generating function $S_4(z)$. To compute this LODE one can apply the Maple package `gfun` [10] or, as in our case, one can use the Mathematica package `GeneratingFunctions` [11]. Note that the only required information is the leading coefficient

$$z-2z^2-z^3-2z^4+z^5 \tag{7}$$

of the LODE. It is well-known that each singularity of $S_4(z)$ is a root of this polynomial. The smallest non-zero root is $\alpha := \frac{1}{2}\left(3-\sqrt{5}\right)$. Thus, the singularity of $S_4(z)$ that is closest to the origin has absolute value at least α. This implies $s_4(N) = \mathrm{O}((1/\alpha+\varepsilon)^N)$, where $1/\alpha \approx 2.618$.

The growth of $s_5(N)$ can be estimated analogously. We arrive at an LODE of order seven, whose leading coefficient is in fact z^2 times the polynomial (7). Hence we obtain the same big-Oh estimate as for $s_4(N)$. As we will see below, these coarse estimates are sufficient, because the contributions of $s_4(N)$ and $s_5(N)$ to the growth of $s(N,N)$ are negligible.

We could now try to find analogous estimates for $s_i(N)$, $i = 1,2,3$ in the same way. However, we need better asymptotics than what can be found by the crude method we have applied so far. This requires finer knowledge of the generating functions. Instead of analyzing the generating functions of the three sums separately, we save some work by using the following "combined" recurrence. It is also useful for rapid computation of the dominant part of $s(N,N)$.

Proposition 7. *The sequence*

$$u(N) := 4s_1(N)+2s_2(N)+2s_3(N)$$

satisfies the recurrence

$$-3(N+2)u(N)-2(N+1)u(N+1)+(N+2)u(N+2)=12\,[N=0] \tag{8}$$

with initial values $u(0)=u(1)=0$. *(We use Iverson's bracket notation:* $[\text{true}] = 1$ *and* $[\text{false}]=0$*).*

Proof. Subsequently, let $f_i(N,k)$ be the summand of $s_i(N,k)$ for $1 \le i \le 3$, i.e., $s_i(N) = \sum_{k=1}^{N-2} f_i(N,k)$. Note that for all $N \ge 1$ and all $2 \le k \le N-2$ the summands can be written in the form

$$f_1(N,k) = -\frac{k(-2+2k-N)(-1+2k-N)}{(-1+k)(k-N)N} b(N,k),$$
$$f_2(N,k) = \frac{(-2+2k-N)(-1+2k-N)}{(-1+k)N} b(N,k),$$
$$f_3(N,k) = \frac{(-2+2k-N)(-1+2k-N)(2k-N)(1+2k-N)}{(-1+k)(k-N)(1+k-N)N} b(N,k)$$

where $b(N,K) = \binom{N}{k}\binom{N-k}{k-2}$. For $N \geq 1$ and $2 \leq k \leq N-2$ define $f(N,k) := 4f_1(N,k) + 2f_2(N,k) + 2f_3(N,k)$, i.e.,

$$f(N,k) = \frac{2(-2+2k-N)(-1+2k-N)(k+3k^2-2N-4kN+2N^2)}{(-1+k)(k-N)(1+k-N)N} b(N,k).$$

The main step of the proof is the certificate recurrence

$$g(N,k+1)-g(N,k) = c_0(N)f(N,k)+c_1(N)f(N+1,k)+c_2(N)f(N+2,k) \quad (9)$$

given by

$$c_0(N) := -3N(2+N), \quad c_1(N) := -2N(1+N), \quad c_2(N) := N(2+N)$$

and

$$\begin{aligned} g(N,k) := 4k\Big(&6k^3(2+N) - k^2(42+47N+14N^2) \\ &+ k(42+77N+50N^2+12N^3) - 2(6+15N+16N^2+9N^3+2N^4)\Big)\times \\ &\times b(N,k)\Big/\big((N-k)(N-k+1)(N-k+2)\big), \end{aligned}$$

which can be found with the package `Sigma` or `Zb`. Assuming the correctness of equation (9), we now sum equation (9) over the summation range. Using the fact that $u(N) = \sum_{k=2}^{N-2} f(N,k) + 4(N-1)$, we easily obtain the recurrence (8).

To establish (9), we first tried to express $b(n,k+1)$, which occurs in $g(N,k+1)$, and $b(N+i,k)$, which occurs in $f(N+i,k)$, in terms of $b(N,k)$ times a rational function in N and k. But this approach failed due to pole problems within the summation range $2 \leq k \leq N-2$; e.g., we have $b(N+1,k) = (N+2)/(N-2k+4)b(N,k)$. In order to avoid this, we express all quantities in (9) in terms of $b'(N,k) := \binom{N}{k}\binom{N-k}{k-4}$ by

$$b(N,k) = \frac{(N-2k+4)(N-2k+3)}{(k-3)(k-2)} b'(N,k),$$
$$b(N+1,k) = \frac{(N-2k+4)(N+1)}{(k-3)(k-2)} b'(N,k),$$
$$b(N+2,k) = \frac{(N+1)(N+2)}{(k-3)(k-2)} b'(N,k),$$
$$b(N,k+1) = \frac{(N-2k+4)(N-2k+3)(N-2k+2)(N-2k+1)}{(k-3)(k-2)(k-1)(k+1)} b'(N,k).$$

With simple polynomial arithmetic one can verify that (9) holds for all $N \geq 6$ and $4 \leq k \leq N-2$. Summing (9) over k from 4 to $N-2$ and compensating missing terms produces, together with a rigorous proof, our recurrence (8) for $N \geq 4$. The special cases $N = 0, 1, 2, 3$ can be verified by simple evaluation. □

The recurrence (8) of the sequence $u(N)$ translates into the differential equation

$$-6zU(z) - \left(3z^2 + 2z - 1\right) U'(z) = 12z, \qquad U(0) = 0,$$

for its generating function $U(z)$, again calculated using the `GeneratingFunctions` package. Since $U(z)$ is analytic at zero, it equals the unique analytic solution of this initial value problem. But the latter can be expressed in terms of radicals (e.g., using Mathematica's `DSolve`), yielding

$$U(z) = \frac{9}{8}\frac{\sqrt{1+z}}{\sqrt{1-3z}} + \frac{3}{8}\frac{\sqrt{1-3z}}{\sqrt{1+z}} + \frac{1}{2}\frac{\sqrt{1-3z}}{(1+z)^{3/2}} - 2. \tag{10}$$

This explicit formula shows that $z = 1/3$ is the smallest singularity of $U(z)$. Thus, its coefficients have an exponential growth rate of 3^N, so we are on the right track in our quest to establish (6). Flajolet and Odlyzko's method of singularity analysis [12, 13] shows how to obtain finer information on the growth of the coefficients from the behaviour of the function near the singularity. The latter is given by

$$U(\tfrac{1}{3}z) = \tfrac{3}{4}\sqrt{3}\,(1-z)^{-1/2} + \mathrm{O}(1) \qquad \text{as} \quad z \to 1^{-}. \tag{11}$$

Theorem 8 (Standard function scale [12, 13])**.** *Let $\alpha \notin \{0, -1, -2, \dots\}$. Then the coefficient of z^N in*

$$f(z) = (1-z)^{-\alpha}$$

satisfies

$$[z^N]f(z) = \frac{N^{\alpha-1}}{\Gamma(\alpha)}\left(1 + \mathrm{O}(N^{-1})\right).$$

Theorem 9 (Big-Oh transfer [12, 13])**.** *Assume that $f(z)$ is analytic in a circle with radius greater than 1, slit along the real half line $[1, \infty[$, and satisfies*

$$f(z) = \mathrm{O}((1-z)^{-\alpha}), \qquad z \to 1.$$

Then the coefficient of z^N in $f(z)$ satisfies

$$[z^N]f(z) = \mathrm{O}(N^{\alpha-1}).$$

Applying Theorem 7 and Theorem 8 to (11) yields

$$\begin{aligned} 3^{-N}u(N) &= \tfrac{3}{4}\sqrt{3}\frac{N^{-1/2}}{\Gamma(1/2)}\left(1 + \mathrm{O}(N^{-1})\right) + \mathrm{O}(N^{-1}) \\ &= \tfrac{3}{4}\sqrt{\tfrac{3}{\pi}}N^{-1/2} + \mathrm{O}(N^{-1}), \end{aligned}$$

whence the final result (Theorem 5); recall that we have established $s_i(N) = O((2.7)^N)$, $i = 4, 5$, above.

Note that an asymptotic expansion of the quantity in parentheses in (6), up to arbitrary order, can be easily derived from (10) with Theorem 7 and Theorem 8.

References

[1] T. Schelling, *Models of Segregation*, American Economic Review, 59 (1969), pp. 488–493.

[2] T. Schelling, *On the Ecology of Micromotives*, The Public Interest, 25(1971), pp. 61–98.

[3] T. Schelling, *Dynamic Models of Segregation*, Journal of Mathematical Sociology, 1 (1971), pp. 143–186

[4] J. Epstein, D. Cummings, S. Chakravarty, R. Singha, and D. Burke, *Toward A Containment Strategy for Small Pox Bioterror*, Brookings Institution Press and the MIT Press (2004)

[5] D. Earn, J. Dushoff, and S. Levin, *Ecology and Evolution of the Flu*, Trends in Ecology and Evolution, 17 (2002), pp. 334–340

[6] M. Pollicott and H. Weiss, *The Dynamics of Schelling-Type Segregation Models and a Graph Laplacian Variational Problem*, Advanced in Applied Mathematics, 27 (2001), pp. 17–40.

[7] C. Schneider, *The Summation Package Sigma: Underlying principles and a rhombus tiling application*, Discrete Math. Theor. Comput. Sci., 6 (2004), pp. 365–386.

[8] P. Paule and M. Schorn, *A Mathematica version of Zeilberger's Algorithm for Proving Binomial Coefficient Identities*, J. Symb. Comput., 20 (1995), pp. 673–698.

[9] D. Zeilberger, *The Method of Creative Telescoping*, J. Symbolic Comput., 11 (1991), pp. 195–204.

[10] B. Salvy and P. Zimmermann, *Gfun: A Package for the Manipulation of Generating and Holonomic Functions in One Variable*, ACM Trans. Math. Software, 20 (1994), pp. 163–177.

[11] C. Mallinger, *Algorithmic Manipulations and Transformations of Univariate Holonomic Functions and Sequences*, Master's thesis, RISC, J. Kepler University, August 1996.

[12] P. Flajolet and A. M. Odlyzko, *Singularity Analysis of Generating Functions*, SIAM Journal on Discrete Mathematics, 3 (1990), pp. 216–240.

[13] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, 2005. Book in preparation: Individual chapters are available as INRIA Research Reports 1888, 2026, 2376, 2956, 3162, 4103 and electronically under http://algo.inria.fr/flajolet/Publications/books.html.