A Markov model of urban evolution: Neighborhood change as a complex process

Daniel Silver*1 and Thiago H Silva2

1 Sociology, University of Toronto, Toronto, Canada
2 School of Cities, University of Toronto, Toronto, Canada

*These authors contributed equally to this work.
* dan.silver@utoronto.ca

Abstract

This paper develops a model of socio-spatial neighborhood evolution, using Toronto, Canada as our case study. Building on neighborhood change research and complexity theories of cities, we address three research questions: 1) What are the main types of neighborhoods in the city and how are they organized, both spatially and hierarchically? 2) What are the main patterns of neighborhood change over time, and how are they inflected by space? 3) How are these trends likely to unfold in the future, and how might they change under certain urban planning scenarios? Cluster analysis reveals three major neighborhood types, which we term “creative city,” “suburban,” and “marginalized,” along with sub-types in which various occupational and ethnic groups tend to predominate or mix. Markov models of transition patterns prove to be highly accurate, successfully predicting the final distribution of neighborhood types with an average RMSE of 0.008. Counterfactual scenarios empirically demonstrate urban complexity: small initial changes reverberate throughout the system, and unfold differently depending on their initial geographic distribution. These scenarios show the value of complexity as a framework for interpreting data and guiding scenario-based planning exercises.
Introduction

This paper seeks to build a model of socio-spatial neighborhood evolution, taking account of the fact that cities are complex systems. Using Toronto, Canada as our case study, we join and extend insights from two related literatures that have largely developed in parallel: neighborhood change research and complexity theories of cities. Building on these traditions, we pursue three research questions: 1) What are the main types of neighborhoods in the city and how are they organized, both spatially and hierarchically? 2) What are the main patterns of neighborhood change over time, and how are they inflected by space? 3) How are these trends likely to unfold in the future, and how might they change under certain urban planning scenarios?

To answer these questions, we examine publicly available quinquennial census data from 1996-2016. Cluster analysis reveals three major neighborhood types, which we term “creative city,” “suburban,” and “marginalized,” along with sub-types in which various occupational and ethnic groups tend to predominate or mix. Markov models of transition patterns prove to be highly accurate, successfully predicting the final distribution of neighborhood types with low error – an average RMSE of 0.008. Examining these models shows that the predominant pattern of change is continuity, as most neighborhood types tend to reproduce themselves over time. However, this tendency is reduced in the city’s interstitial areas. Spatial Markov models show more precisely the extent to which neighborhood evolution is influenced by its surrounding areas, indicating a deep interconnection between urban space and time. Counterfactual scenarios empirically demonstrate urban complexity: i) small initial changes in transition probabilities can lead to larger transformations that reverberate throughout the system; ii) the spatial distribution of changes in initial conditions generate divergent evolutionary outcomes.

Beyond the specific findings, our study seeks to advance research in both socio-spatial neighborhood change and urban complexity. Regarding neighborhood change research, we make a number of methodological contributions: whereas most such research relies on either top-down or bottom-up clustering methods, we take a hybrid multilevel approach. Instead of sequence analysis, we use Markov models to identify the generative processes that produce patterns of change, exploiting the formal properties of
Markov chains to do so. In addition, we add a spatial component to the Markov process. Moreover, in contrast to most past work in this area, we validate the predictive power of the model and, having done so, demonstrate how to use such models to evaluate hypothetical scenarios to evaluate how the city would evolve under different conditions. Regarding complexity theories of city, our primary contribution is to demonstrate the utility of core concepts in an empirical setting. If much past work in this tradition has been in a mathematical and simulation context, we bring complexity ideas – recursivity, threshold effects, non-linearity, and more – into dialogue with empirical neighborhood research and show its value as a heuristic for interpreting data and guiding scenario-based planning.

We proceed in 4 sections. First, we review related work on neighborhood change and urban complexity, while also providing relevant background about Toronto. Second, we discuss data and methodology. Third, we report results, organized around the three research questions noted above. We conclude by discussing our results, noting limitations, and considering future directions.

Related work

This study joins and builds upon two research streams: socio-spatial neighborhood change and complexity theories of cities. In this section, we review key concepts and trends in both, note how our study advances them, and articulate our research questions that emerge from them. We also provide a brief overview of relevant work about our research setting, Toronto, Canada.

Socio-spatial neighborhood change research

The tradition of socio-spatial neighborhood change research extends back roughly a century, to the Chicago and Atlanta schools of urban sociology 1,2. These early efforts pioneered methods of close observation and data collection, including contributing to the modernization of census bureaus and the creation of standardized census tract boundaries to facilitate neighborhood comparison 3. Even if they have been criticized and modified over time, many of the key concepts and typologies from this period remain relevant and in circulation today 4. These include the concentric zone model,
in which groups compete for prime central space and change radiates out from there or the neighborhood succession and filtering model in which neighborhoods defined by upper status groups are replaced in cycles by lower status groups.

Mid-century researchers carried this tradition forward with the aid of novel statistical techniques. Shevsky and Bell in particular elaborated factorial ecology as a research paradigm for multivariate classification of neighbourhood types and change. Factorial ecology utilized factor analysis to reduce many variables to a small number of components, which in turn were used to characterize neighborhoods and their changes (see and ). This research suggests that neighborhoods could be differentiated along three primary dimensions, which summarize their economic, family, and ethnic status. While contemporary research often adds additional factors such as housing, transit, and built form, a key legacy of factorial ecology is to incorporate variables describing these dimensions.

If the multidimensional character of factorial ecology never entirely disappeared from academic research, for several decades it fell by the wayside. Instead, neighborhood and urban change research tended to focus on single variables, often in the form of segregation indices. Nevertheless, largely in market research, geodemographics emerged in this period and promoted cluster analysis as an alternative approach to detecting neighborhood types and change as complex combinations of many variables. In particular, geodemographic researchers argued that cluster analysis could detect both major types and important but rare types that tend to be lost in factor analysis.

In recent decades, these geodemographic techniques began to appear more in an academic context. A key advance was to perform the cluster analysis on many variables across multiple time points, and then to compare transitions over time across the resulting neighborhood types. While these studies failed to exploit the analytical potentials in the underlying transition matrices, they did generate some important results, primarily with regard to US neighborhoods: Stability (including persistent poverty) is much more common than change. Polarization over time (declining middle class neighborhoods) is a general tendency. Neighborhood succession (downgrading) does not necessarily follow the linear trajectory envisioned by earlier theorists, and varies greatly by city. Upgrading (including gentrification) occurs less
frequently than one might imagine from popular discussions and is focalized in ethnically
diverse neighborhoods located in a few large metro areas. Finally, this research
emphasized the existence of a great diversity of neighborhood change types, many of
which do not neatly fit into the categories of stability, succession, or upgrading in that
they involve movement among neighborhood types with similar socioeconomic status.

More recent studies of socio-spatial neighborhood change have advanced in formal
 sophistication as well as geographic and temporal scope. The work of Delmelle
(e.g. [28]) has moved the field forward decisively in these regards. Perhaps most
importantly, this has occurred by making neighborhoods’ longer-term temporal
trajectories the primary unit of analysis and adapting sequence analysis techniques from
genomics (see also [29]). This work largely confirmed key substantive results from
previous US studies regarding the main patterns of change. However, it added by
locating temporal pathways in space, finding for instance that high poverty minority
and wealthy white neighborhoods expand in a relatively contiguous pattern whereas
multiethnic areas grow in a more dispersed pattern.

Our study continues and extends these past related works. Like most recent work,
we use cluster analysis to classify neighborhood types and examine transitions through
those types. We differ by using a method that joins both hierarchical and
prototype-based methods (hierarchical k-means, discussed below) and incorporating the
hierarchical (nested) structure of the resulting neighborhood typology into the analysis.
This approach enables us to meaningfully include more clusters than is typical in the
literature. Similarly, we carry forward the tradition of studying transition probabilities
and examining spatial patterns of change, but add by mapping transition rates
themselves (rather than sequences). Likewise, we examine the impact of spatial context
through spatial Markov chains. Perhaps most significantly, in contrast to past work we
exploit the formal properties of Markov models to evaluate their predictive power based
on real data and also considering the effects of various counterfactual scenarios. This is
done on the spatial and non-spatial cases. Here we highlight concepts such as tipping
points and non-linearity that are characteristic of complexity theories of cities.
Cities as complex systems

While socio-spatial neighborhood change research has often been very empirically focused, especially through its historical legacy to the Chicago School of urban sociology, it has maintained a theoretical basis that points toward key concepts from complexity thinking. Abbott \[30\] has stressed this theoretical linkage, noting the importance in this tradition of “turning points” (similar to non-linearity and tipping points), fractal patterns, the emergence of global order from local interaction, and the deep integration of space and time \[30\].

Many of the insights driving efforts to conceive cities as complex systems derive from the work of Jane Jacobs and Christopher Alexander in the 1960s. The core ideas researchers took from these authors included the notion that behind the seeming disorder and diversity of urban forms there are strong patterns, and these emerge from the myriad local decisions and behaviors by individual actors and organizations \[31\]. These processes of self-organization suggested treating cities as complex systems. They exhibit emergent orders that require significant energy to sustain, patterns of segregation rooted in competition over space, and punctuated equilibria \[32\]. In particular, complexity studies of cities have tended to emphasize how self-organization involves three major steps \[33\]: emergence (where local interactions between parts give rise to complex global structures); steady states (where the system maintains itself); and bifurcation or phase transition (in which the old equilibrium dissolves and a new one emerges).

The major contribution of complexity theories of cities has been primarily theoretical. Researchers have sought to bring diverse urban phenomena, often considered independently, into a common framework \[33\]. Much work in particular has sought to elaborate the mathematical basis for this synthesis. These have joined with considerable use of computer simulation models, especially Agent Based Models and Cellular Automata \[31,34\], to demonstrate self-organization and emergence in general and highlight some of their specific urban patterns (such as larger zones of stability and local persistent areas of experimentation and change at their boundaries \[35\]). At the same time, the sorts of long-term empirical studies of local and multilevel change characteristic of socio-spatial neighborhood change research are relatively rare (though...
While much of this work unfolded in a mode of rapid theoretical and methodological growth, more recent studies start from the point of view that complexity theories of cities have “come of age” \[37\]. They are a relatively mature set of theories and methods, and the main challenges involve clarification, application, and refinement. An area of particular concern involves demonstrating the relevance of complexity ideas to substantive areas of urban studies to avoid the fate of past quantitative research paradigms, which in pursuing narrow technical questions lost sight of larger concerns \[38\]. Heavy reliance on simulations geared toward short-term predictions similarly makes considering longer-term transformations difficult, leading some researchers to develop more explicitly evolutionary models \[39\].

A related concern involves challenges in importing assumptions from physics into urban studies, given that cities are composed of human agents that are themselves highly complex cognitive-emotional entities that react to changes in their environments \[40\]. This has made incorporating complexity theories into urban planning programs difficult, in that complexity stresses a certain degree of unpredictability, non-linearity, and multiple forking paths that are not easily assimilated into clear planning interventions on the model of an engineer pulling levers in a machine \[41\]. While in general complexity theories caution humility in planning, they also seek to rethink the nature of planning theory to some degree. In particular, some advocate scenario based thinking that considers possible futures under different assumptions, while incorporating sensitivity to the notion that small initial differences can compound into large changes over time \[37,42\].

Our study draws from complexity theories of cities primarily as a heuristic guide toward interpreting the results of our analyses and formulating our questions. More specifically, considering cities as complex systems leads us to treat neighborhood change as a multilevel process in which transitions are constrained by a vertical hierarchy but also occur horizontally across higher-order boundaries. Similarly, we seek to locate persistent patterns of stability and volatility. Perhaps most importantly, we adopt the scenario-based approach as a way to imagine possible urban futures. Here we formulate our scenarios in the light of complexity theory by considering hypothetical changes to the city’s transition probabilities strategically located at the boundaries among its
higher order structuring principles. Crucial as well is our emphasis on how small changes can have large consequences that reverberate throughout the system, affecting other parts indirectly.

Synthesizing and extending this work, this study pursues three interrelated research questions to uncover:

1. The main types of neighborhoods and their organization (spatially and hierarchically);

2. The key patterns of neighborhood change over time, and how space affects them; and

3. How these trends are likely to unfold in the future, including under different urban scenarios.

**Contextualizing the case of Toronto**

Our study uses Toronto, Canada as a case study for examining socio-spatial neighborhood change within a complex urban setting. To contextualize our findings, we provide some brief background information about recent socio-spatial trends in Toronto.

Especially since the early 1990s, Toronto has experienced dramatic social and economic changes along several fronts, which have in turn manifested in its social geography. The city has undergone a rapid shift to a post-industrial economy: manufacturing jobs, mostly located in the inner suburban zone, have declined, service and finance have increased, and the post-industrial “creative occupations” from fine artists to graphic designers to technology and R&D have rapidly grown, mostly in the downtown core [43]. This led to a long-term economic decline of the suburban zone relative to the gentrifying core, marked by a hollowing out of middle class neighborhoods in the former and a growth of young professional and “creative class” enclaves in the latter [44,45].

At the same time, a new wave of immigrants have settled in Toronto. Previous immigrants had primarily hailed from Southern and Eastern Europe in the 1950s and 1960s, while more recent cohorts have come from East and South Asia, Africa, the Caribbean, and the Middle East in steeply rising proportions and absolute numbers through the 1980s and into the 21st century [46]. Most recent immigrants have settled
in the city’s suburban zone and in its aging social housing high-rise towers (built in the 1950s-1970s), which now contain some of the most ethnically-diverse neighborhoods in the world. This has led to the emergence of lower income suburban ethnic neighborhoods predominated by newer immigrant groups, often alongside or nearby older Southern and Eastern European communities. These are generally located in the city’s Northeast and Northwest areas where rapid transit service is relatively poor, whereas older established upper status suburban neighborhoods generally follow the subway lines.

Population growth in the past several decades has been rapid. In 2019, Toronto was the fastest growing city in North America, both in the city proper and the wider metro area – gaining nearly as many new residents as New York lost. Concurrently, average household size has been declining, as more persons are living alone, with roommates or partners, or in small nuclear families, often in dense clusters of single family dwellings or newly built downtown condominiums, which have appeared in rapidly increasing numbers. In 2019, Toronto had the most construction cranes of any city in North America, more than New York, Seattle, Boston, and Los Angeles combined. The bulk of this activity is located in the downtown core, though new condominiums are also appearing in central districts in the city’s suburban zones, continuing its distinctive polycentric pattern of development.

Still, even amidst these dramatic and rapid transformations, a fundamental three-part structure has persisted. First, there is a dense downtown core that is highly walkable and well-served by multiple forms of transit, where the built form is predominantly office buildings and newly built condominiums amidst tree-lined pre-war single-family home neighborhoods. This area is largely populated by young people and highly educated “creative class” professionals and aspirants, punctuated by social housing towers with many lower income and immigrant residents. The other two parts of the city are located in its suburban zone. In its traditional suburban areas one finds large single family homes often inhabited by members of Canada’s historical Protestant elite working in managerial occupations, though nearby are Toronto’s main Jewish areas as well as areas settled by newer Chinese immigrants. Third, there are the city’s relatively marginalized inner suburban areas in the Northeast and Northwest. These are marked by a mix of single family homes and social housing towers, populated primarily
by lower income non-white residents (in particular South Asian, Black, and Arab) along with older white, often Italian, working class communities. In this core-periphery pattern, Toronto resembles other large post-industrial cities such as London, UK [50].

Data and methodology

Data sources and preprocessing

We use publicly available data from 1996-2016 for census tracts located within the Toronto Census District (CD) [51] representing five different points in time $t \in T = \{1,2,3,4,5\}$. Consistent with past studies [9, 24–27], we classify neighborhoods according to a broad set of demographic, socioeconomic, and housing characteristics (all variables used is in Table 1).

**Table 1. Variables used in the analysis.**

<table>
<thead>
<tr>
<th>Variable Type</th>
<th>Variable Label</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socioeconomic</td>
<td>marmarried</td>
<td>% married</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>unemp</td>
<td>% unemployed</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>labmanag</td>
<td>% management occupations</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>labartsport</td>
<td>% arts, recreation, and sports occupations</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>labbusi</td>
<td>% business, finance, and administrative occupations</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>labsci</td>
<td>% natural and applied science occupations</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>labserv</td>
<td>% sales and service occupations</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>labblue</td>
<td>% blue collar occupations (trades + manufacturing)</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>commauto</td>
<td>% drive to work</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>incavhh</td>
<td>average household income</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>educba</td>
<td>% with BA degree or higher</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>incgov</td>
<td>% of income from government transfer</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>popdens</td>
<td>population density</td>
</tr>
<tr>
<td>Demographic</td>
<td>ageyouth</td>
<td>% population 25-34</td>
</tr>
<tr>
<td>Demographic</td>
<td>agesenior</td>
<td>% population 65 over</td>
</tr>
<tr>
<td>Demographic</td>
<td>vmblack</td>
<td>% black visible minority</td>
</tr>
<tr>
<td>Demographic</td>
<td>vm.sasia</td>
<td>% south asian visible minority</td>
</tr>
<tr>
<td>Demographic</td>
<td>vm.chi</td>
<td>% chinese visible minority</td>
</tr>
<tr>
<td>Demographic</td>
<td>vm.arab</td>
<td>% arab visible minority</td>
</tr>
<tr>
<td>Demographic</td>
<td>eth.ital</td>
<td>% italian ethnicity</td>
</tr>
<tr>
<td>Demographic</td>
<td>eth.port</td>
<td>% portuguese ethnicity</td>
</tr>
<tr>
<td>Demographic</td>
<td>eth.grk</td>
<td>% greek ethnicity</td>
</tr>
<tr>
<td>Demographic</td>
<td>eth.jew</td>
<td>% jewish ethnicity</td>
</tr>
<tr>
<td>Housing</td>
<td>dwdetach</td>
<td>% detached housing</td>
</tr>
<tr>
<td>Housing</td>
<td>dwhirise</td>
<td>% apartments over 5 stories</td>
</tr>
<tr>
<td>Housing</td>
<td>tenurerenters</td>
<td>% renters</td>
</tr>
<tr>
<td>Housing</td>
<td>dwellval</td>
<td>average dwelling value</td>
</tr>
<tr>
<td>Housing</td>
<td>move5</td>
<td>% moved in past 5 years</td>
</tr>
</tbody>
</table>

As census-tracts boundaries can change over time, a fundamental challenge when making a longitudinal comparison of spatial data is to ensure that boundaries are consistent. Our dataset minimizes this problem by utilizing the methodology proposed by Allen and Taylor [52]. This approach harmonizes census-tracts to a common set of boundaries, using map-matching, dasymmetric overlays, and population-weighted

July 11, 2020
interpolation. Some census-tracts did not contain data for some variables, and for the purposes of this analysis they were removed from the dataset (2.5% of all tracts). In total, we study 514 census-tracts: $c_i \in C = \{c_1, c_2, \ldots, c_{514}\}$. We represent each census-tract on the n-dimensional space $V^{c.t} = \{v_1^{c.t}, v_2^{c.t}, \ldots, v_n^{c.t}\}$ defined by all the $n = 28$ census variables. Next, all variables are normalized by z-score each year to control for different measurement scales and to enable a fair comparison of changes across time. In other words, we create a new vector $F^{c.t} = \{f_1^{c.t}, f_2^{c.t}, \ldots, f_n^{c.t}\}$, where each variable $f_i^{c.t} = (v_i^{c.t} - \mu_c^{c.t})/\sigma_c^{c.t}$, with $\mu_c^{c.t}$ and $\sigma_c^{c.t}$ being the average and standard deviation of $V^{c.t}$, respectively. Thus, $F^{c.t}$ represents a specific tract $c$ on the census year $t$. In this way, the normalized variables in $F$, therefore, represent a relative value compared to all other values in the city in a particular year.

**Overall workflow**

Our workflow is in part inspired by the work of Elizabeth Delmelle [28], regarding the way clustering is used to extract socioeconomic typologies, though it is consistent with other work in the field [28]. Fig 1 summarizes our main methodological steps, explained in the next sections.

**Fig 1. Overall workflow.**

**Socioeconomic typology and temporal mapping**

In order to create a longitudinal socioeconomic typology for the census-tracts under evaluation, we perform a clustering on all census-tracts $c \in C$ represented by $F^{c.t}$ observed on the five years studied. Most neighborhood classification studies use either a prototype-based clustering (typically k-means) or hierarchical clustering [18], with ongoing debates about their relative merits [25]. In our study, we strike a middle ground by combining both approaches. We use a k-means clustering algorithm to partition our data. However, instead of randomly choosing the initial centroids, which makes the results sensitive to these choices, we select these centroids based on hierarchical clustering (using Euclidean distance with Ward linkage criteria) [53]. This approach is also known as hierarchical k-means. This hybrid method has the benefit of producing a stable solution in a computationally efficient way, while also allowing us to examine the
general organization of the clusters within the resulting hierarchical data space. This in turn can also guide decisions about the number of clusters to use in the analysis based on a dendrogram.

Following the clustering procedure, we interpret the results according to procedures typical in the neighborhood classification literature. We use radar plots to identify each cluster’s semantic based on their centers’ most prominent variables. Each cluster then receives a unique name \( u \in U \) that tries to express its latent meaning. While these names are helpful in making sense of the analysis, their subjective nature does not compromise the results – using different names as labels could provide a similar message. Each tract is then assigned one of those labels five times (each one representing a point in time corresponding to a census year). Finally, a sequence for each tract is created, depicting its longitudinal trajectory through the multidimensional neighborhood typology: \( \text{longitudinal trajectory } \tau_i = (u_{t=1}, u_{t=2}, u_{t=3}, u_{t=4}, u_{t=5}) \), for the five census years under study. We also employ t-Distributed Stochastic Neighbor Embedding (t-SNE), which is a tool for dimensionality reduction that is well suited for the visualization of high-dimensional data, as in our case [54]. Specifically, we use t-SNE to identify similar neighborhood types based on the clusters’ centroids and to place them in a higher-order grouping (in this study, the three most distinct sets of neighborhood types).

**Markov models**

Having generated a longitudinal trajectory for each census-tracts through the neighborhood types, we create two types of Markov chain: a first order Markov chain and a spatial Markov chain. These tools enable us to study the evolution of neighborhood types over time.

**First order Markov chain**

Markov chains are stochastic processes that can be parameterized by empirically estimating transition probabilities between discrete states [55]. A random variable \( X \) represents a sequence of states, where a discrete level on \( X \) is the state of this variable in a certain time. The Markov chain of the first order, explored in this study, is one for
which the probability of $X$ being in state $j$ (neighborhood type, in our case) at time $t$ depends only on the state immediately preceding one $i$ of $X$, at time $t - 1$. For $K$ states, the first order transition matrix has a size of $K \times K$, where $K$ is the size of the number of unique neighborhoods types ($|U|$) and takes the form:

$$
\begin{bmatrix}
m_{1,1} & m_{1,2} & \ldots & m_{1,K} \\
m_{2,1} & m_{2,2} & \ldots & m_{2,K} \\
\vdots & \vdots & \ddots & \vdots \\
m_{K,1} & m_{K,2} & \ldots & m_{K,K}
\end{bmatrix}
$$

(1)

The probabilities from transitioning from state $i$ to $j$ on $M$ can be estimated from the relative frequencies of the transitions on all longitudinal trajectories for all tracts. Thus, each transition probability is:

$$m_{ij} = \frac{\sum_{t=1}^{T-1} a_{i',j+1}/t}{\sum_{t=1}^{T-1} a_{i,t}}$$

(2)

where $a_{i',j+1}$ is the amount of neighborhoods transitioning from type $i$ in year $t$ to type $j$ in $t + 1$, and $a_{i,t}$ is the total amount of neighborhoods in type $i$ in period $t$.

Before examining this Markov chain, we validate its performance as a predictive model of the city’s evolution. For that, we derive a Markov chain from years 1-4 (1996-2011). Thus, in this Markov chain $t \in T' = \{1, 2, 3, 4\}$, and, consequently, the longitudinal trajectory for each tract is represented by the sequence: $(u_{t=1}, u_{t=2}, u_{t=3}, u_{t=4})$. We use that new Markov chain to predict the fifth year (2016).

For that, we use the probabilities of having each neighborhood type in one specific year in the past $\pi(t) = [\pi_1(t), \pi_2(t), \ldots, \pi_K(t)]$, where $i$-th element $\pi_i(t)$ represents the probability of a neighborhood type in year $t \in \{1, 2, 3, 4\}$, to predict $\pi(t = 5)$. Thus, we make four distinct predictions for ($t = 5$), each of them considering $\pi(t)$ for one of the previous years studied.

We compare the model’s predicted proportions of neighborhood types to the actual proportions in year five, against two baseline models: a) a $K \times K$ identity matrix, b) 1000 randomly generated first order transition matrices $M_x$, where $x \in \{1, 2, \ldots, 1000\}$. 

Each $M_x$ is derived from a random order of the set of all longitudinal trajectory tracts, by randomly choosing four neighborhood types in $U$ to create a new sequence for each tract studied: $(u_x^{t=1}, u_x^{t=2}, u_x^{t=3}, u_x^{t=4})$. We explore $M_x$ to perform predictions about the distribution of neighborhood types on year 5, $\pi(t=5)$, using for that $\pi(t=1)$, $\pi(t=2)$, $\pi(t=3)$, and $\pi(t=4)$ on each $M_x$. With the identity matrix we disregard the Markov chain, and our prediction of year $\pi(t=5)$ is $\pi(t)$, with $t \in \{1, 2, 3, 4\}$.

We explore also $M$ to perform predictions about the distribution of neighborhood types 50 time steps in the future. To do so, we use $\pi(t=5)$, to make predictions based on $M$. This results in estimations of probabilities of having each neighborhood type in Toronto in the future.

**Spatial Markov chain**

Spatial Markov chains allow for a more comprehensive study of the spatial influence of the transitional dynamics [56]. In this study, we investigate whether transition probabilities are dependent on neighborhood types. Instead of estimating one transition probability matrix, as done in the non-spatial case, spatial Markov chains require estimation of $K$ transition probability matrices, each of which is conditional on the neighborhood type at the year immediately before. A key step in the implementation of a spatial Markov chain is the definition of a neighborhood’s vicinity. Here we include the five nearest neighbors, and we calculate conditional probabilities based on those neighbors. A spatial Markov matrix decomposes the non-spatial $K \times K$ transition matrix into a $K \times K \times K$ system, where $K$ is the number of neighborhood types under study. Thus, we have one transition matrix $K \times K$ for each type of neighborhood, and each of them represents transition probabilities conditioned on neighboring a specific neighborhood type, considering the year immediately before. We represent a spatial Markov for each neighborhood type $u \in U$ as $S^u$. If a certain row of $S^u$ sums to zero, occurring when a certain neighborhood type was never a neighbor of the specific one it is being conditioned on, we assign 1 to the position representing the diagonal element to ensure each conditional transition probability matrix, i.e., $S^u$, is a valid stochastic matrix (each row sums up to 1) [56]. In other words, the probability of staying with the same neighborhood type is 1 in that specific conditional case.

We use $S^u$ to perform predictions about the state of the city 50 steps in the future.
To do so, we again use the probabilities of having each neighborhood type in year five, $\pi(t = 5)$, in this case to make predictions based on $S^w$. This results in estimates of the probabilities of having each neighborhood types in Toronto in the future, $P^w$,

conditioned on a certain type of neighbor $u \in U$.

We also perform a test of spatial independence, comparing the transition probabilities conditioned on neighboring a specific neighborhood type (spatial case) with transition probabilities considering the entire dataset without this restriction (non-spatial case). For that, the transition probabilities under $H_0 : \forall u : p_{ij|u} = p_{ij|u}(u = 1, 2, \ldots, U)$ are contrasted against those under $H_a : \exists u : p_{ij|u} \neq p_{ij}$, using a Pearson $\chi^2$ test statistic as presented by Bickenbach and Bode [57]:

\[
Q(M) = \sum_{u=1}^{[U]} \sum_{i=1}^{[U]} \sum_{j\in A_i} p_{ij|u} \frac{(\hat{p}_{ij|u} - \hat{p}_{ij})}{\hat{p}_{ij}} \sim asy \chi^2 \left( \sum_{i=1}^{[U]} (a_i - 1)(b_i - 1) \right)
\]

(3)

where $A_{ij|u} = j : \hat{p}_{ij|u} > 0$ is the set of nonzero transition probabilities in the $i$-th row of the transition matrix estimated from the $u$-th neighbor type, $A_i = j : \hat{p}_{ij} > 0$ is the set of nonzero transition probabilities in the $i$-th row of the transition matrix estimated from the entire dataset (non-spatial), $a_i = \#A_i$ is the number of elements in $A_i$, and $b_i = \#B_i$ is the number of submatrices (for the spatial case) for which a positive number of observations is available for the $i$-th row [55].

**Counterfactual scenarios**

We also create counterfactuals scenarios to demonstrate the utility of our approach to, for example, guiding urban planning decisions. Thinking in counterfactuals demands envisioning a hypothetical reality that contradicts the known facts, namely census tract data in our case. Here we evaluate the impact of imaginative interventions in neighborhoods, by either changing their transition probabilities or changing their initial conditions.
Non-spatial case

Following ideas from complexity theories of cities, we explore the extent to which small initial changes, when repeatedly iterated, can lead to relatively large and sometimes unexpected changes, both direct and indirect. Specifically, we consider two scenarios. On the first one, Scenario 1, we induce small increases, 2% each, to the probability that three neighborhood types \( UP = \{ \text{"black predominant"}, \text{"mixed suburban"}, \text{"mixed creative"} \} \) – would appear in three entrenched neighborhood types – \( DOWN = \{ \text{"elite suburban"}, \text{"established creative"}, \text{"young urban professional"} \} \) – decreasing their reproduction rates correspondingly by 6% each:

\[
\sum_{i=1}^{3} \sum_{j=1}^{3} (UP_i \rightarrow DOWN_j) + 0.02 \quad \text{and} \quad \sum_{i=1}^{3} (DOWN_i \rightarrow DOWN_i) - 0.06,
\]

where \( X \rightarrow Y \), represents a specific transition on the Markov chain \( M \). In this way, we ensure a valid Markov model in this hypothetical scenario. This scenario represents a strategic planning decision to promote interchange among parts of the city that rarely interact and to induce change in some of the city’s most entrenched upper status areas (where reproduction rates are near or above .9). Indeed, transitions between these neighborhood types are exceedingly rare: all are below 3% and most are near 0. In the second scenario, Scenario 2, we increase the same transitional probabilities by 1% instead of 2% (and correspondingly reduce the reproduction rate of the target neighborhood types by 3% each). Comparing these two scenarios allows us to investigate threshold effects.

Spatial Case

For this experiment, we imagined a scenario where an intervention in the city changed certain neighborhoods of the type “towers” to “mixed suburban” (the one that was found to have the biggest spatial influence, as shown in the analysis section).

Considering all neighborhood types of all longitudinal trajectories, we consider two different scenarios. In Scenario 1, we randomly select five neighborhoods if they are classified as “towers” in the fourth year, i.e., \( u_{t=4} \), and change them to “mixed suburban” in year five, \( u_{t=5} = \text{"mixed suburban"} \), for all selected neighborhoods. In Scenario 2, we randomly select one tract classified as “towers” in year four, and the immediate four nearest neighbors of the same type also in year 4, and change all of
them to “mixed suburban” in year five. While the non-spatial scenarios allow us to examine threshold effects, these spatial scenarios investigate geographical distributional effects. Moreover, these scenarios allow us to examine how small changes in the initial distribution of neighborhood types affect the long term evolution of the city, even if its transition probabilities are not altered.

With these adjustments in the longitudinal trajectories for both scenarios, we create a new spatial Markov chains $S'^{u'}$, following the same steps for the spatial (non-counterfactual) case. We also perform predictions about the distribution of neighborhood types 50 time steps in the future, obtaining $P'^u$. To do so, we use $\pi(t = 5)$ to make predictions based on $S'^u$. This whole process is repeated 100 times, and we work with their average values: $\frac{1}{100} \sum_{i=1}^{100} P'^{u'}_i$ for every neighborhood type $u \in U$.

**Results**

Our analysis proceeds according to the three research questions articulated above.

**What are the main types of neighborhoods in the city and how are they organized, spatially and hierarchically?**

A challenge in any clustering task is the number of groups (or clusters) to feature, and there is no simple solution. Many neighborhood clustering studies use various statistical techniques as guides, such as within-cluster sum of squares or the silhouette method. In neighborhood research, these tend to suggest a relatively small number of clusters, often between 4 and 7. For this study, we chose to highlight a somewhat larger number, beginning our search in the 10-15 range. This allows us to retain the benefit of multi-dimensional clustering – the ability to identify latent meanings arising from combinations of factors – while remaining closer to the underlying variables. With a small number of clusters, changes in these variables may not appear.

To select the specific number of clusters for our analysis, we examine the dendrogram represented in Fig 2.

**Fig 2. Dendrogram of clustering results.**

Looking at the clusters in the 10-15 range (height around 50), the dendrogram
suggests focusing on 13 or 14 clusters. We qualitatively examined both of these solutions, and found that the 14 cluster solution did not add significant analytical value in the present context and so opted for the more parsimonious 13 cluster solution. In the 14 cluster solution, the predominantly Asian portions of the city’s inner suburbs, which are largely geographically contiguous, are further subdivided, mostly by socio-economic status. Examining neighborhoods at this level of specificity could be valuable for further research focused on this segment of the city. At the same time, we situate the 13 clusters within the context of the city’s higher-order structure, three larger groups, as suggested by the dendrogram. Particularly, for this interpretation task we use t-SNE (as mentioned in the methodology section), which confirmed the distinct separability of three groups (see S1 Fig).

Table 1 summarizes the neighborhood typology that emerges from the clustering process. Radar plots with complete results are in the Supporting Information (S2 Fig).

Table 2. Hierarchical typology of Toronto neighborhoods.

<table>
<thead>
<tr>
<th>Creative City</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>arts &amp; culture, science and tech, education, density, high rises, youth, renters, movers.</td>
<td>drivers, detached housing, unemployment, married couples, black, south asian, service workers, blue collar</td>
</tr>
</tbody>
</table>

**Established Creative, Mixed Creative, Young Urban Professionals, Portuguese Predominant**

<table>
<thead>
<tr>
<th>Suburban</th>
<th>Suburban</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>managers, business administration, science and tech, drivers, income, education, home value, detached housing, seniors, married couples</td>
<td>service workers, blue collar, government income, density, high rises, youth, renters, movers, black, south asian, unemployment</td>
</tr>
</tbody>
</table>

**Elite Suburban, Mixed Suburban, Chinese Predominant, Jewish Predominant, Greek Predominant**

<table>
<thead>
<tr>
<th>Marginalized</th>
<th>Marginalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>service workers, blue collar, government income, renters, black, south asian, unemployment, high rises</td>
<td>managers, arts &amp; culture, business admin, science &amp; tech, income, education, home value, detached housing</td>
</tr>
</tbody>
</table>

**Towers, Black Predominant, South Asian Predominant, Italian Predominant**

This table summarizes the multilevel neighborhood typology that emerges from our cluster analysis. The typology divides the city into three higher order types of neighborhoods, “Creative City,” “Suburban,” and “Marginalized,” which are characterized primarily by high and low values on the variables listed below each label. As Fig 3 shows, these types also exhibit a higher degree of similarity. More specific neighborhood types that fall within each of these higher-order classifications are listed below them (in bold).

Overall, Toronto is characterized by three main types of neighborhoods with various...
subtypes:

1. “Creative city” neighborhoods tend to be located in the downtown core and resemble the types of neighborhoods featured in Richard Florida’s “The Rise of the Creative Class” \[58\]: high density areas populated by mobile, highly educated young people, singles, artists, science and tech workers, with relatively few low income blue collar and service workers. More specific types include “established creative” neighborhoods with very high concentrations of arts culture workers, high home values, single family homes, few blue collar and service workers, and relatively low density; “mixed creative” areas, located at the fringes of the downtown core, that exhibit more occupational and ethno-racial diversity; and “young urban professional” neighborhoods featuring very high concentrations of highly educated young professional singles focalized in the city center. Toronto’s predominantly Portuguese neighborhoods continue to have a foothold in this part of the city.

2. “Suburban” neighborhoods by contrast are high income, low density areas with fewer youth and more homeowners, high home values, much detached housing, and numerous managers and businesspeople. This part of the city is largely made up of neighborhoods defined by socio-economic status and ethnicity. “Elite” suburban neighborhoods have the city’s highest incomes, most managers, lowest density, highest home values, and fewest blue collar and service workers, while “mixed” suburban areas have middle incomes and greater occupational and ethnic diversity. Other suburban neighborhoods are defined more by a predominant ethnic community: Greek, Jewish, or Chinese.

3. The city’s “marginalized” areas, also located generally outside the downtown core, tend to be characterized by lower income non-white residents, higher unemployment, more blue collar and service workers, lower home values, and high-rise towers. This part of the city contains dense, high-rise social housing complexes –“towers”– marked by high unemployment, high turnover, and concentrations of diverse non-white immigrant communities (especially Arab, South Asian, and Black residents). Other working class neighborhoods in this part of the city are also ethnically diverse but also have a relatively predominant ethnic
group: Black, South Asian, or Italian.

This typology resembles other neighborhood classifications of Toronto derived by different methods \[44,50\]. These also identify three major clusters of neighborhoods with similar geographical, socioeconomic, and demographic patterns. Our typology adds specificity by identifying subtypes within these major patterns and allowing us to examine the complex multilevel interactions between the higher-order structure of the city and its lower-level constituents.

What are the main patterns of neighborhood change over time, and how are they inflected by space?

Having identified the structure and semantic of the city’s neighborhoods, we turn to our second research question. To answer this question, we first generate a first order Markov chain describing the transition matrix between neighborhood types and validate its performance as described in the methodology. As Table 2 shows, the model is highly accurate, with consistently lower RMSE values than the baseline models (the only exception is predictions based on the identity matrix from year 4). Having validated the model, we generate a Markov chain from the complete dataset (years 1-5) and use that to examine the city’s patterns of change.

Table 3. Evaluating the error of the model.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Real</th>
<th>Identity</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t = 1$</td>
<td>0.00848</td>
<td>0.02313</td>
<td>0.11657</td>
</tr>
<tr>
<td>$t = 2$</td>
<td>0.01016</td>
<td>0.01519</td>
<td>0.11573</td>
</tr>
<tr>
<td>$t = 3$</td>
<td>0.01063</td>
<td>0.01261</td>
<td>0.11556</td>
</tr>
<tr>
<td>$t = 4$</td>
<td>0.00762</td>
<td>0.00571</td>
<td>0.11569</td>
</tr>
</tbody>
</table>

Values in the table represent a RMSE from all predictions exploring different predictors: $t = 1$ is the state on year one, i.e., $\pi(t = 1)$; $t = 2$ is the state on year two, $\pi(t = 2)$; $t = 3$ is the state on year three, $\pi(t = 3)$; and $t = 4$ is the state on year four, $\pi(t = 4)$. For the random experiment, all confidence interval values for each predicted neighborhood type were below 0.001, i.e., the predictions obtained do not change significantly, and they were omitted to favor legibility. Real refers to the results with real data, Identity to the results with an identity matrix, and Random to the results for the random experiment.

Fig 3 depicts the transition matrix as a network (the full matrix is in the Supporting Information: S3 Fig). As in previous neighborhood change studies, the dominant pattern is continuity. Most of the time, neighborhoods remain similar from one year to
the next: all have at least a .7 probability of reproducing themselves, and most (9/13) are above .85. In this way, each new iteration of the city is at once the same and subtly different from its predecessor. Dramatic change is rare, but for that reason it stands out all the more when it occurs. This helps to explain why the predictions based on the identity matrix considering year 4 to predict year 5 have a good performance.

**Fig 3. Network representation of neighborhood transition probabilities.** This figure summarizes the transition probabilities across neighborhood types as a simplified network. To increase legibility, the figure excludes transitions with a probability less than or equal to .02 and only labels probabilities greater than or equal to .05 on edges. Edge widths are proportional to transition probabilities (excluding self-loops) and nodes are colored to correspond to the higher-order structures of which each neighborhood is a member. Node sizes are proportional to the number of census tracts of that type.

The most stable types of neighborhoods are “elite suburban” and “young urban professional,” while the least stable are mixed (suburban and creative) or European ethnic neighborhoods (Greek, Portuguese). These latter tend to transition into either “mixed” or “established creative” neighborhoods, while “mixed creative” neighborhoods in turn are relatively unlikely to sustain themselves. Instead they are highly likely (.15) to transition into “established creative” types. If we think of this pathway as gentrification, then the most likely targets are older European immigrant neighborhoods. Nevertheless, in virtue of their somewhat interstitial position (they are located in the Western and Eastern transitional zones between the “creative city,” “suburban,” and “marginalized areas”) these neighborhoods are subject to a diversity of transitions. For example, Greek predominant neighborhoods tend to transition into “South Asian predominant,” “Chinese predominant,” and “established creative” areas – spanning all three higher-order types. By contrast, “mixed suburban” neighborhoods, which also show a high degree of volatility, tend to transition into newer immigrant areas, which show little propensity to gentrify. This overall pattern whereby middle class suburban areas are diminishing while high status suburban and downtown youth areas persist matches other recent studies of Toronto. At the same time, the prevalence of horizontal transitions – “Chinese predominant” with “mixed suburban,” “towers” with “black predominant” illustrate the diversity of forms of change identified in past US studies.

While the city exhibits a high degree of continuity, its degree of structure itself
varies. Rather than assume there is an equally powerful structure at work throughout, an important question concerns how much of a given urban environment is structured at all, and to what degree. We find that much of Toronto is very deeply structured, so that it is highly likely to reproduce itself. In fact, some of the most “creative” parts of the city in terms of who is there and what they are doing – areas in which young, highly educated arts and technology workers predominante – are the most stable. By contrast, other parts of the city exhibit creativity where the urban fabric itself is in a state of transition in which neighborhood forms themselves rise and fall more rapidly. There are therefore at least two types of urban creativity at work here. One appears to thrive within a stable urban context that supports a specific set of groups and activities; in the other, the urban form itself is in a more fluid state of experiment and transformation.

These results reveal a complex hierarchical structure governing urban evolution. The major ordering principles of the city reproduce themselves and maintain the differences between them, even as there is a relatively large amount of movement to and from different local formations within them. The transition probabilities for the higher order principles (creative city, suburbs, marginality) are all above .9. Moreover, as Fig 4 shows, the greatest stability through time tends to be located at the geographic center of these areas. At the same time, amidst these “high fidelity” zones, there are smaller zones of volatility. They are generally located in interstitial areas near the edges of the three higher-order zones, that is, where “creative city” borders “marginalized” and “suburban” areas or “marginalized and “suburban” border one another. This spatio-temporal pattern reveals an important dimension of urban complexity that has been highlighted in computer simulation studies of the self-organization of cities: zones of local instability within a more global spatio-temporal stability, where the areas “in between - the boundaries - are thus the most critical areas in the city for socio-spatial changes (and in society at large).” This description of an artificially generated city is remarkably similar to the one we have drawn from real-world data on Toronto.

Fig 4. Map of neighborhood reproduction rates. This figure maps the reproduction rates of neighborhoods across the city. “Reproduction rates” refer to the probability that a given neighborhood type will recur, that is, the diagonal value in the transition matrix. In this map, each census tract is assigned its neighborhood type’s diagonal value, according to the thirteen neighborhood types listed in Table 1. Pink areas are more volatile, yellow areas are more stable. Tract borders are colored to indicate each tract’s position in the higher-order typology depicted in Table 1.

July 11, 2020
To investigate these spatio-temporal dependencies more precisely, we generate a spatial Markov chain according to the methodology described above. Full results are shown in the Supporting Information. Before we proceed, we evaluate the influence of neighborhood type on transition probabilities on immediate spatial surroundings (our spatial case). For that we use a test of spatial independence, as described in the methodology section. Our results show that spatial dependence in terms of surrounding neighborhood type has a statistically significant impact on transition probabilities. According to the $\chi^2$ test, the system is not independent across space because the test statistic is $Q = 2354.1$ ($p < 0.001$, DoF = 854), thus, rejecting the null hypothesis of spatial independence. This means that a state’s transition cannot be assumed to be independent from the neighborhood type of its neighbors.

After that, we highlight some overall patterns and particularly interesting findings in Fig 5. Given the high degree of spatial clustering, many neighborhood types are never or rarely nearby and therefore the number of possible spatially dependent transitions is often low. We therefore restrict the analysis to transitions that occur at least 10 times, causing an increase or decrease of at least 5% on the reproduction rate (diagonal value) compared to the non-spatial case.

**Fig 5. Spatial markov model results regarding differences on reproduction rates.** This figure represents an increase (blue) or decrease (red) of at least 5% on the reproduction rate (diagonal value) on the spatial case compared to the non-spatial case. Numbers on top and right are the total counts of colored cells for each column and line, respectively.

Five key results are depicted in Figure 5:

- First, spatial dependence is high in general. Across most neighborhood types the reproduction rate shifts substantially depending on its neighbors. This was the case for 10/13 neighborhood types, as we can observe by looking at the columns of Fig 5. For instance, the neighborhood type “mixed creative” is influenced in this regard by six different types of neighbor.

- Second, some neighborhood types show more spatial dependence than others. Types “south asian predominant” and “mixed creative” show the greatest tendency to become more volatile in the presence of neighbors, while “young urban professionals”, “elite suburban”, and “jewish predominant” neighborhood
types are less sensitive to their surrounding geographic context (their reproduction rates do not significantly change depending on nearby neighborhood types).

- Third, some neighborhood types induce greater increases or decreases in the reproduction rates of their neighbors. “Jewish predominant”, “Italian predominant”, “elite suburban”, “Portuguese predominant”, “established creative”, and “mixed creative” induce significant decreases in reproduction rates of their neighbors (red colors in Fig 5); by contrast, the remaining neighborhood types favor increasing stability in their surroundings (blue color in Fig 5), for instance, “towers” enhances the stability of “mixed creative” and “south asian prevalent”.

- Fourth, beside diagonals, some particular transitions are especially affected by their spatial location. In particular, the transition probability “Portuguese predominant” → “mixed creative” shifts from 0.103 to 0.188 when near “mixed creative” and the transition probability ”mixed creative” → ”established creative” shifts from 0.146 to 0.205 when near ”established creative.” These are situations where neighborhood change exhibits spatial tipping, in which the appearance of a neighboring type harkens impending nearby transformations, in particular in these cases toward the spread of a specific type (“mixed creative” or “established creative”) to its neighbors.

- Finally, the city’s multilevel neighborhood order affects the spatial dependencies of its temporal evolution, though this impact varies. In general, “marginalized” neighborhood types become more volatile when they are near “creative city” neighborhoods, but the reverse happens less frequently. This adds more specificity to the results depicted above in Fig 4.

Altogether, these findings provide strong support for the view of urban evolution advanced by complexity theories of cities. Simultaneous stability and openness arises from a confluence of top-down and bottom-up processes. Zones of local volatility are globally stable and patterns of global change contain local stability. The temporality of urban evolution is grounded in space, just as the spatial arrangement of the city emerges from its neighborhoods’ divergent trajectories through time.
How are these trends likely to unfold in the future, and how might they change under certain urban planning scenarios?

To address our third and final research question, we use the Markov chains described above to predict how the city is likely to evolve in the future, and how this evolution would change under different scenarios. Fig 6 shows the difference between the predicted distribution of neighborhood types and the actual 2016 distribution in 50 time steps, along with two counterfactual scenarios described in the methodology section.

Fig 6. Predicted changes in neighborhood distribution according to three models. This figure summarizes predicted changes in the overall distribution of neighborhood types in 50 time steps for three models. Blue squares show the non-counterfactual scenario in which the city continues to evolve according to its current pattern. Red circles and green triangles show the counterfactual scenarios described in the methodology section.

The non-counterfactual results (blue squares) in Fig 6 show how the city would evolve if it continued forward according to its current trajectory. The dominant trend would be increased socio-economic polarization: elite suburban neighborhoods would increase their share of the overall population of neighborhoods by around 7.5 percentage points (from 8 to 15.5%), and “young urban professional” and “established creative” areas by around 5 percentage points (from 11 and 7 percent, respectively). All together, these three upper status areas would grow from around 26 to 35% of the city as a whole. At the same time, middle income diverse suburban neighborhoods would decline (by about 5 percentage points), along with most of the city’s ethnic working and service class communities, as well as its “mixed creative” neighborhoods. If unchecked, current trends point toward a solidification of the “divided city” [60].

A benefit of our methodological approach is that it allows us to envision alternative scenarios that might change this trajectory towards polarization and division. Three key points stand out in examining the counterfactual scenarios in Fig 6. First, in line with complexity theories, small initial changes can have big effects. In both scenarios, the growth of “young urban professional,” “elite suburban,” and “established creative” areas is substantially reduced. By contrast, the decline in the city’s occupationally and ethnically diverse areas is reduced or stabilized in “mixed creative,” “Chinese predominant,” “portuguese predominant,” and “south asian predominant” areas. In
some cases, such as predominantly black neighborhoods, the trend reverses to net
growth. Second, we see some signs of threshold effects or phase shifts, again in line with
complexity theories of cities. The incremental change from Scenario 1 (.01 change in
transition probabilities) to Scenario 2 (.02 change in transition probabilities) generates
relatively sharp downstream effects, most strikingly in the case of “young urban
professional,” “elite suburban,” and “black predominant” neighborhood types. And
third, we see evidence of indirect effects characteristic of complex systems: while we did
not make any change to the transition probabilities for “south asian predominant” or
“tower” neighborhoods, their relative footprint in the city grew compared to the
non-counterfactual scenario.

All in all, these results show that in a complex dynamic interacting system, small
quantitative changes at critical points can potentially make a substantial qualitative
difference. Connecting disconnected and divided upper status areas with lower status
areas reduces the isolation of these parts of the city, and helps others to retain their
foothold. This in turns reveals another sign of a complex system: changes in one part
reverberate in others.

Fig 7 summarizes the results of the spatial counterfactual scenarios described in the
methodology. Here we consider the effects of changing 5 tracts from “towers” to “mixed
suburban,” while leaving other transition probabilities unchanged. In Scenario 1, the 5
tracts are randomly distributed around the city; in Scenario 2, they are geographically
clustered. These situations allow us to examine the impacts of the initial distribution of
neighborhood types on the evolution of their geographic arrangement. For both
scenarios, the figure shows where there is a difference of at least 1% in relation to the
non-counterfactual case when predicting the distribution of neighborhood types. More
specifically, if in the non-counterfactual case the model predicts a given neighborhood
type will constitute 10% of the total tracts nearby the target type, and in the
counterfactual case the distribution is greater than 11% or less than 9%, we color the
appropriate box in the matrix: blue (at least 1% increase) and red (at least 1%
decrease). Rows indicate the conditioned neighborhood type on the spatial Markov over
all other ones (columns).

Fig 7 indicates a substantial similarity between the two scenarios, as we would
expect. The median value regarding the distribution of all differences between the same
Fig 7. Results of counterfactual scenarios for the spatial case. This figure summarizes the spatial counterfactual scenarios’ results (Scenarios 1 and 2), showing where there is a difference of at least 1% in relation to the non-counterfactual case when predicting the distribution of neighborhood types: blue (at least 1% increase) and red (at least 1% decrease). Labels on the y-axis left indicate the conditioned neighborhood type on the spatial Markov over all other ones (x-axis bottom). Numbers on top and right are the total counts of colored cells for each column and line, respectively.

transition probabilities in Scenario 1 and Scenario 2 is 0, with a standard deviation of 0.0066. Also, there is no difference bigger than 3% compared to the non-counterfactual case in any scenario. Nevertheless, some nuances emerge on closer inspection. For instance, while the growth of “mixed suburban” in both scenarios is highly similar, Scenario 1 (Fig 7-left) induces declining shares of “towers” and “black predominant” neighborhoods in the geographic context of “south asian predominant”, with all the averages statistically different under a 95% confidence interval. In addition, the reproduction rate of “black predominant” is higher on Scenario 1 (average of 0.192 with +−0.008 95% C.I.) compared to Scenario 2 (average of 0.206 with +−0.002 95% C.I.).

In both scenarios, we note a slight tendency toward a potential form of gentrification. For example, we observe a considerable increase in the probability of a neighborhood of type “established creative” having a neighbor of type “mixed creative.”

These results suggest that introducing a small change in the initial spatial distribution of neighborhood types can lead to overall urban transformation, and this tends to be enhanced if these changes are more widely spread than concentrated in one geographic area.

Discussion and conclusion

Our paper has sought to advance the study of socio-spatial neighborhood change and cities as complex systems, substantively, methodologically, theoretically, and practically. Substantively, we have described a distinctive multilevel neighborhood classification in which lower-level neighborhood types defined predominantly by socioeconomic features and the physical environment are members of more global structures. In turn, we have shown that major temporal patterns of change through these classifications exhibit distinctive spatial structure whereby changes are focalized at interstitial zones in the typology. Finally, we have shown that a city with these current types of neighborhoods
and evolutionary tendencies is likely to show sharply increasing polarization over time if these patterns continue unabated. And we have envisioned scenarios in which small changes to the current transition probabilities or distributions of neighborhood types could alter this trajectory.

Methodologically, this study makes several contributions. Primarily, we show the potential of mixed clustering methods for socio-spatial neighborhood change research, and the analytical utility of exploiting the formal features of Markov chains in general and spatial Markov chains in particular. We elaborate how to use these models in imaginative exercises geared toward envisioning how changing the present might alter the future in expected and unexpected ways. Though the primary ambition of this study is not theoretical, in advancing dialogue between complexity theories of cities and neighborhood change research, we lay the ground for more fulsome efforts at theoretical synthesis. Finally, the study suggests ways to incorporate complexity concepts into planning practice. In particular, it shows how to utilize scenario-based thinking to envision the potential consequences of hypothetical interventions. While we have highlighted a few scenarios of possible strategic value, the methods can be used to creatively explore other cases and scenarios. Encouraging this experimental approach to city life is a key contribution of the framework proposed here.

While this study has made a number of significant contributions, it is not without limitations. These point toward important areas of future research. One limitation concerns temporal and geographic scope. Carrying these methods further into the past and into comparative studies of other cities and countries would greatly enhance their value. Doing so is challenging because of data comparability issues. Recent advances have helped to overcome the challenge of changing definitions of neighborhood borders but changing variable definitions (i.e. in occupational or ethnic labels) remain a major obstacle. Solving this problem would enable longer-term studies at a fine-grained level. Another limitation concerns the use of first-order Markov chains. These by definition do not consider memory of past transitions, which in the urban context is likely considerable. Future research, especially of longer-term trajectories, would benefit from incorporating higher-order Markov chains. Despite obtaining strong evidence that there is spatial dependence in the neighborhood socioeconomic evolution, which could be helpful in many ways, this study would also benefit from further
research trying to empirically explain and prove causation regarding those patterns. Finally, while we have pursued a Markovian approach oriented toward generative processes rather than overall sequences, synthesizing these two approaches rather than placing them in competition with one another strikes us as an important way to move the field forward, both theoretically and empirically.

Supporting information

S1 Fig. Visualization of high-dimensional clusters’ centroids representing neighborhood types in 2-dimensions. This figure represents the result of t-SNE on the clusters’ centroids of neighborhood types. It is possible to identify three distinct groups.

S2 Fig. Radar plots of clusters. These are radar plots summarizing the cluster means of 13 neighborhood types listed in Table 1. They show values for each neighborhood type for the variables listed in Table 1 (the suffix “.s” is added to indicate these are standardized).

S3 Fig. Non-spatial and spatial markov chains. Heat maps representing the non-spatial markov chain and all the thirteen spatial markov chains, according to the names indicated on each figure. For instance, the figure with the title “Spatial - Neighboring Mixed Creative” represents the spatial Markov chain whose transitions are conditioned to the neighborhood type “Mixed Creative”.

Acknowledgments

We are grateful for ongoing discussion of these issues examined in this paper with members of the Urban Genome Project, in particular Scott Sanner, Mark Fox, Rob Wright, and Fernando Calderon-Figueroa.

References


35. Portugali J. Self-Organization and the City. Haken H, editor. Springer Series in
Synergetics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2000. Available from:
http://link.springer.com/10.1007/978-3-662-04099-7

36. Pumain D. Urban systems dynamics, urban growth and scaling laws: The
question of ergodicity. In: Complexity theories of cities have come of age. Berlin,

37. Tan E, Portugali J. The responsive city design game. In Complexity Theories of

Theory of Suburbanization: Capitalism and the Construction of Urban Space in
the United States,'in Dear and Scott; p. 383–430.

39. Yamu C. It Is Simply Complex(ity): Modeling and Simulation in the Light of
Decision-Making, Emergent Structures and a World of Non-Linearity. disP - The

40. Zamenopoulos T. A complexity theory of design intentionality. Artificial
Intelligence for Engineering Design, Analysis and Manufacturing.

Meyer H, Stolk E, Tan E, editors. Complexity Theories of Cities Have Come of
from: http://link.springer.com/10.1007/978-3-642-24544-2_11

42. Sijmons D. Simple Rules: Emerging Order? A Designer’s Curiosity About
Complexity Theories of Cities Have Come of Age. Berlin, Heidelberg: Springer
Berlin Heidelberg; 2012. p. 281–309. Available from:
http://link.springer.com/10.1007/978-3-642-24544-2_16

43. of Toronto C. Toronto employment survey 2013. Toronto: City of Toronto; 2014.

44. Hulchanski JD, University of Toronto, Cities Centre. The three cities within


60. Florida R. The divided city: And the shape of the new metropolis. Martin Prosperity Institute; 2014.
Multidimensional Census Data

Clustering (Socioeconomic typology creation)

Temporal Mapping (Longitudinal sequences for each area)

First Order Markov Chain Creation

Spatial Markov Chain Creation

Predictions (Real and counterfactual scenarios)
<table>
<thead>
<tr>
<th></th>
<th>Mixed Creative</th>
<th>Established Creative</th>
<th>Portuguese Predominant</th>
<th>Young Urban Prof</th>
<th>Towers</th>
<th>Greek Predominant</th>
<th>Mixed Suburban</th>
<th>Elite Suburban</th>
<th>Italian Predominant</th>
<th>Jewish Predominant</th>
<th>Black Predominant</th>
<th>Chinese Predominant</th>
<th>South Asian Predominant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed Creative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Established Creative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portuguese Predominant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young Urban Prof</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Towers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greek Predominant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed Suburban</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elite Suburban</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian Predominant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jewish Predominant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black Predominant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese Predominant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Asian Predominant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig 6: Change over time for different demographic groups.
Click here to access/download
Supporting Information
S3_fig.eps