Interim and Long-Run Dynamics in the Evolution of Conventions☆

David K. Levine[1], Salvatore Modica[2]

Abstract

We characterize transitions between stochastically stable states and relative ergodic probabilities in the theory of the evolution of conventions. We give an application to the fall of hegemonies in the evolutionary theory of institutions and conflict and illustrate the theory with the fall of the Qing Dynasty and rise of Communism in China.
JEL Classification Numbers: C73 (Evolutionary Games)

Keywords: Evolution, Conventions, Markov Chains, State Power

☆First Version: October 4, 2013. We are grateful to NSF Grant SES-08-51315 and to the MIUR PRIN 20103S5RN3 for financial support.

*Corresponding author David K. Levine, Villa San Paolo, Via della Piazzuola 43, I-50133 Firenze - Italy

Email addresses: david@dklevine.com (David K. Levine), salvatore.modica@unipa.it (Salvatore Modica)

[1]Department of Economics, European University Institute and Washington University in St. Louis

[2]University of Palermo, Italy

1. Introduction

The modern theory of the evolution of conventions deals with a Markov process in which there are strong forces such as learning towards "equilibrium" and weak "evolutionary" forces such as mutations that disturb an equilibrium and lead from one equilibrium to another. To prove theorems, the limit as the weak forces approach zero is analyzed. In the limit itself equilibria appear as irreducible classes of the Markov process: sets of states all of which are accessible to each other, but grouped into classes which are isolated from each other. Near the limit - the situation of interest - the Markov process is ergodic, and puts positive weight on all states. However some states are more equal than others, and in the unique limit of the stationary distributions weight is placed only on the irreducible classes of the limit - and moreover, only some of these classes have positive weight - the stochastically stable classes. In particular, while the limiting Markov process can have many equilibria, the limit of the Markov processes may place weight on only one or a few equilibria. The literature, especially Kandori, Mailath and Rob (1993), Young (1993) and Ellison (2000), develops a set of techniques for determining which of these equilibria get weight in the limit, and gives a useful picture of what the stochastic process looks like when the weak forces are small but not zero. Roughly, one observes the equilibria that have positive weight in the limit most of the time, but there will inevitably be interruptions in which the system moves temporarily out of equilibrium and back again, and also transitions in which movement takes place from one equilibrium to another.

The focus of the existing literature has been on determining which equilibria get weight in the limit - and this is important to understand. But the transitions - the movement from one equilibrium to another - are also interesting and important for economics. For example: in a model of evolution such as that of Levine and Modica (2013) where different economic and political institutions compete with each other the irreducible classes correspond to hegemonies - a single society that controls all economic resources - and the stochastically stable classes are the most powerful hegemonies. There is considerable historical evidence for the existence of hegemonies: China, the Roman Empire and so forth. In the theory - as in reality - these hegemonies inevitably fall. How the fall takes place - the transition - is of some interest. Do the eventual winners of the conflict appear on the scene and battle back and forth with the hegemony for a while until they take over and establish their own hegemony (short answer - no) or does something else happen, and if so what?

The mathematical methods used in analyzing stochastic stability contains clues for what the transitions might look like. In particular, stochastically stable classes can be characterized by trees of irreducible classes where the distance between classes is measured by "resistance" and the stochastically stable classes appear as the root of least resistance trees. Because of the role played by least resistant paths in this analysis, a natural conjecture is that least resistant paths are in some sense more likely than higher resistant paths. The goal of this paper is to establish in exactly what sense this is true.

The starting point is to observe that a basic feature of resistance is that if we compare the probability of two paths when evolutionary forces are very weak, the lower resistance path is far more likely than the higher resistance path. If we are interested, however, in all the paths between one irreducible class and another, though, the problem is that there are typically many more high resistance paths than low resistance paths. For example, if there is resistance to an invader gaining a piece of land, there are many paths in which the land is lost and regained - which has fairly high resistance, but very few in which it is taken only once and not subsequently lost. Moreover, there is a sense in which over a longer time horizon the relatively more numerous are the high resistant paths than the low resistance paths. If the transition takes quickly, there is not much opportunity for dawdling around losing and regaining land. If it takes place slowly then there are many possibilities for doing so.

Consider then more specifically the set of "direct" paths that lead from one irreducible class to another - without, however, passing through a third class.[3] When evolutionary forces are weak least resistant paths are far less likely than higher resistant paths. The reason for this is simple: it is likely to take a very long time to reach another irreducible class - in the meantime there are likely to be many failed attempts to get there, and these attempts will typically involve some resistance. On the other hand, if we look at the set of paths that have least resistance only after the path leaves the irreducible class for the final time, the transition to the other class is likely to happen relatively quickly and we are able to show that these paths - which describe the transition process itself - are far more likely than other paths.

We establish the theory in two parts. We first develop a set of bounds for direct paths and then for quasi-direct paths, which may linger in an irreducible class for some time before moving on. To illustrate the theory, we apply it to a simplified version of the Levine and Modica (2013) evolutionary model of conflict and the emergence of hegemonies - some details of which are motivated by the transition theory of Acemoglu and Robinson (2001) - illustrating the theory with an account of the fall of the (last) Qing dynasty in China.

Finally, understanding transitions gives us clues about the ergodic probabilities. By examining which irreducible classes are reached "next" from a given starting point we construct a straightforward recursive algorithm that gives precise bounds on the ratio between the ergodic probabilities of all states that are "reasonably close" to irreducible classes.

2. An Illustrative Example

We are interested in economic models that can be represented as Markov processes where some transitions are much less likely than others. To illustrate this we start with an example of a "standard" evolutionary model. Consider the 2x2 symmetric coordination game with actions G, B and payoff matrix

[3] From Ellison (2000) we know that such paths are not necessarily the quickest way of getting to the target, an issue we carefully account for.

	G	B
G	$2,2$	$0,0$
B	$0,0$	$1,1$

This game has two pure Nash equilibria at GG and BB and a mixed equilibrium with probability $1/3$ of G.

To put this in an evolutionary context, we assume that there are five players and that the state of the system is the number of players playing G, so that the state space Z has $N = 6$ states. Each period one player is chosen at random to make a move. We first define the "behavior rule" or "deterministic dynamic" representing "rational" learning: the player whose gets to move chooses a best response to the actions chosen by the opposing players. In addition there are independent trembles: with probability $1-\epsilon$ the behavior rule is followed, while with probability ϵ the player's choice is uniform and random over all possible actions. The presumption is that the chance of "arational" play ϵ is small compared to the probability $1-\epsilon$ of "rational" play.

This dynamic can be represented as a Markov process on the state space Z defined above with six states representing the number of players playing G. Denoting source states by rows and target states by columns as is standard in the theory of Markov chains, the transition matrix can be computed as

$$P_\epsilon = \begin{pmatrix} 1-\frac{\epsilon}{2} & \frac{\epsilon}{2} & 0 & 0 & 0 & 0 \\ \left(\frac{1}{5}-\frac{\epsilon}{10}\right) & \left(\frac{4}{5}-\frac{3\epsilon}{10}\right) & \frac{4\epsilon}{10} & 0 & 0 & 0 \\ 0 & \left(\frac{2}{5}-\frac{2\epsilon}{10}\right) & \frac{\epsilon}{2} & \left(\frac{3}{5}-\frac{3\epsilon}{10}\right) & 0 & 0 \\ 0 & 0 & \frac{3\epsilon}{10} & \left(\frac{3}{5}-\frac{\epsilon}{10}\right) & \left(\frac{2}{5}-\frac{2\epsilon}{10}\right) & 0 \\ 0 & 0 & 0 & \frac{4\epsilon}{10} & \left(\frac{4}{5}-\frac{3\epsilon}{10}\right) & \left(\frac{1}{5}-\frac{\epsilon}{10}\right) \\ 0 & 0 & 0 & 0 & \frac{\epsilon}{2} & 1-\frac{\epsilon}{2} \end{pmatrix}$$

The situation then is as follows. When $\epsilon = 0$ there are two irreducible classes consisting of the singleton sets $\{0\}, \{5\}$ - these sets are each absorbing. We denote the set of irreducible classes corresponding to $\epsilon = 0$ and the transition matrix P_0 by $\Omega = \{\{0\}, \{5\}\}$. They are the pure Nash equilibria of the game. The set of points for which the probability of eventually reaching $\{0\}$ is one - the basin of $\{0\}$ - consists of the points $\{0\}, \{1\}$. The basin of $\{5\}$ is $\{3\}, \{4\}, \{5\}$. The state $\{2\}$ is in the "outer basin" of both $\{0\}$ and $\{5\}$: it has positive probability of reaching either of the two irreducible classes.

We are interested, however, not in $\epsilon = 0$ but in ϵ positive but small. In this case we can talk about what happens "typically" or "most of the time" meaning in the limit as $\epsilon \to 0$. From the earlier results of Young (1993) we know that the system will spend most of its time at $\{5\}$.[4]

[4] Because this system has a special property (the radius of $\{5\}$ greater than the co-radius) the waiting times are also known from Ellison (2000). The waiting time from $\{0\}$ to $\{5\}$ is roughly ϵ^{-2} while the waiting time from $\{5\}$ to $\{0\}$ is roughly ϵ^{-3}. In addition, because the process is a special kind of chain known as a birth-death process, the ergodic distribution may be computed directly to derive some of the following results.

When applied to this example the results of the present paper will give the following additional information on the dynamics of the system:

1. When the transition from $\{0\}$ to $\{5\}$ takes place typically once the state $\{2\}$ is reached, there is no return to the state $\{1\}$ and the transition is very fast.
2. When the transition from $\{5\}$ to $\{0\}$ takes place typically the states $\{4\},\{3\}$ are reached in that order and once the state $\{3\}$ is reached there is no return to the state $\{4\}$ and once $\{2\}$ is reached there is no return to the state $\{3\}$ and the transition is very fast.
3. Starting at $\{0\}$, before $\{5\}$ is reached the system will spend most of the time at $\{0\}$ but will many times reach the state $\{1\}$ for brief periods
4. Starting at $\{5\}$, before the state $\{0\}$ is reached the system will spend most of the time at $\{5\}$ but will many times reach the states $\{4\},\{3\}$ for brief periods
5. The state $\{4\}$ will occur roughly as often as the state $\{0\}$ but while $\{0\}$ will be seen for long stretches of time, the state $\{4\}$ will be seen frequently but only briefly before reverting to $\{5\}$.

3. The Model and the Result for Direct Routes

In the general case we are given a finite state space Z with N elements and a family P_ϵ of Markov chains on Z indexed by $0 \leq \epsilon < 1$. This family satisfies two regularity conditions:

1. $\lim_{\epsilon \to 0} P_\epsilon = P_0$
2. there exists a resistance function $0 \leq r(x,z) \leq \infty$ and constants $0 < C \leq 1 \leq D < \infty$ such that $C\epsilon^{r(x,z)} \leq P_\epsilon(z|x) \leq D\epsilon^{r(x,z)}$

In the illustrative example, the resistances can be computed from the Markov transition matrix as

$$r = \begin{pmatrix} 0 & 1 & \infty & \infty & \infty & \infty \\ 0 & 0 & 1 & \infty & \infty & \infty \\ \infty & 0 & 1 & 0 & \infty & \infty \\ \infty & \infty & 1 & 0 & 0 & \infty \\ \infty & \infty & \infty & 1 & 0 & 0 \\ \infty & \infty & \infty & \infty & 1 & 0 \end{pmatrix}$$

As in the example, we let Ω be the union of the irreducible classes of P_0. We write $\Omega(x)$ for the irreducible class containing x where $\Omega(x) = \emptyset$ if x is not part of an irreducible class. A path a is a finite sequence $(z_0, z_1, \ldots, z_t)$ of points in Z and we write $t(a)$ for the number of transitions in the path. The resistance of the path $r(z_0, z_1, \ldots, z_t) \equiv r(z_0, z_1) + r(z_1, z_2) + \ldots + r(z_{t-1}, z_t)$ with the convention that for the *trivial path* with $t(a) = 0$ then $r(a) = 0$.

We summarize some well known properties of P_0 and Ω. Non-empty irreducible classes $\Omega(x) \neq \emptyset$ are characterized by the property that from any point $y \in \Omega(x)$ there is a positive probability path to any other point $z \in \Omega(x)$ and that every positive probability path starting at y must lie entirely within $\Omega(x)$. Since positive probability in P_0 is the same as zero resistance, we may equally say that from any point $y \in \Omega(x)$ there is a zero resistance path to any other point $z \in \Omega(x)$ and that every zero resistance path starting at y must lie entirely within $\Omega(x)$.

An additional useful notion is this: define a set W to be *comprehensive* if for any point $z \in Z$ there is a positive probability (zero resistance) path to some point in W. In particular the set Ω is comprehensive. We can give the following characterization of a comprehensive set:

Proposition 1. *A set W is comprehensive if and only if it contains at least one point from every non-empty irreducible class.*

Proof. Sufficiency: for any point $z \in Z$ there must be a zero resistance path to some point y in some irreducible class $\Omega(y)$. By assumption there must be a point $w \in W \cap \Omega(y)$, and there is a zero resistance path from any point in $\Omega(y)$ to any other, hence from y to w. Hence the path from z to y and continuing on to w has zero resistance. Necessity: if there is a set $\Omega(y) \neq \emptyset$ with $\Omega(y) \cap W = \emptyset$ then the zero resistance path to W assumption fails: any zero resistance path originating in $\Omega(y)$ must remain entirely within $\Omega(y)$ and hence does not reach W. □

3.1. The Concept of Direct Routes

We now define a *forbidden set W* for a path a to be a set that the path does not touch except possibly at the beginning and end. Given an initial point $x \in Z$ with $\Omega(x) \neq \emptyset$ and sets $W \subseteq Z$ and $B \subseteq W$, we call a non-trivial path from x to B with forbidden set W a *direct route* if W is comprehensive and the path has positive probability for $\epsilon > 0$. Then for each x, B and comprehensive W there is a set $\mathcal{A}(x, B, W)$ of direct routes from x to B with forbidden set W. [5]

To motivate these definitions and our subsequent results, consider the problem of moving - not necessarily directly - from one non-empty irreducible class $\Omega(x)$ to a different non-empty irreducible class $\Omega(y)$. If $\epsilon = 0$ this is impossible. For $\epsilon > 0$ it may be possible. However, in order to leave $\Omega(x)$ to get to $\Omega(y)$ at some point the path must leave $\Omega(x)$ and then hit some point in Ω, say a point in $\Omega(z)$ (where it may be that $z = y$). That is, at some point, there must be a direct route from some point $x' \in \Omega(x)$ to some set $B = \Omega(z)$ with forbidden set $W = \Omega$. The fact that $W = \Omega$ just reflects the idea that after leaving x' there is no return to $\Omega(x)$ until $\Omega(z)$ is reached, and that the first point in Ω that is reached is a point in $\Omega(z)$. Such paths are improbable, yet they are important as they are needed to move from one irreducible class to another.

An intuition for why direct routes have particular stochastic properties is this. In P_0 a path that hits a point in an irreducible class is then trapped in that class, so cannot reach a target outside of that class. As we noted, when $\epsilon > 0$ this need not be the case. However, if a point in an irreducible class is hit then it is very likely that the path will then linger in that irreducible class passing through every point in the class many times. Hence there is a sense in which paths that do not hit a comprehensive set are "quick" - they cannot linger in an irreducible class for if they did so they would have to hit every point in the class many times, thus touching W. By contrast if W were not comprehensive, then paths not passing through W could linger in an irreducible

[5] The assumption that $B \subseteq W$ is without loss of generality. We can always define a forbidden set $W' = W \cup B$ without changing the set of direct routes.

set for a long time. Direct routes are a bit like the hare in the story of the tortoise and the hare. Direct routes get to the destination quickly - they must if they are not to fall into the forbidden set. Because of this, as Ellison (2000) points out, they are not very reliable: routes that linger in an irreducible class may be far more likely than direct routes to reach their destination.

We are interested in the following questions: how likely is the set of direct routes $\mathcal{A}(x, B, W)$, which paths in $\mathcal{A}(x, B, W)$ are most likely, what are these paths like and how long are they? There are two uses of these results: first, they are useful tools that we will use to analyze interim dynamics in subsequent sections. Second, they provide insight into interim dynamics in some special cases.

To illustrate the usefulness of the concept of direct routes and its limits in analyzing interim dynamics we consider two examples where x is a singleton irreducible class, that is $\Omega(x) = \{x\}$. In the first we take $B = \Omega \backslash \Omega(x)$ and $W = \Omega$; then the direct routes are the paths that leave x and hit a different irreducible class without returning to x. If we think of irreducible classes as "equilibria" then $\mathcal{A}(x, B, W)$ represents transitions paths from one equilibrium to some another. Hence we are asking how likely is it that we go to another equilibrium, which are the most likely routes to another equilibrium, and how long does it take to get there? In the other example $y \notin \Omega(x), \Omega(y) \neq \emptyset$, and $B = \Omega(y), W = \Omega$; then $\mathcal{A}(x, B, W)$ is the set of the paths that leave x and hit $\Omega(y)$ without first returning to x or hitting any other irreducible class. Now suppose we are interested in the question: what are all the possible routes (not necessarily direct) from x to B which do not return to x before hitting B. In the former case all such routes are direct routes - hence analysis of direct routes is sufficient. In the second case an analysis of direct routes is not sufficient because there are routes that are not direct - those which hit some point of $\Omega \setminus \Omega(y) \subseteq W$ before first hitting B.

There are two other important examples involving $\{x\} = \Omega(x) \neq \emptyset$. First take what Ellison (2000) calls the *basin* of $\Omega(x)$, that is the set of points $\mathcal{B}(x)$ in Z for which there is a zero resistance path to $\Omega(x)$ and no zero resistance paths to $\Omega \backslash \Omega(x)$. Or what amounts to the same thing, the set of points for which there is probability one in P_0 of returning to $\Omega(x)$. We may then consider $B = Z \backslash \mathcal{B}(x), W = B \cup \{x\}$. In this case $\mathcal{A}(x, B, W)$ are the paths that leave the basin of $\Omega(x)$ without first returning to x. We can also define the *outer basin* of $\Omega(x)$ as the set of points $\mathcal{B}^+(x)$ for which there a zero resistance path back to $\Omega(x)$. In this case $Z \backslash \mathcal{B}^+(x)$ represents the basins of irreducible classes other than $\Omega(x)$. Then we may be interested in $B = Z \backslash \mathcal{B}^+(x), W = B \cup \{x\}$. In this case $\mathcal{A}(x, B, W)$ are the paths that hit the basins of other irreducible classes without first returning to x.[6]

3.2. Results on Direct Routes

The intuition behind the results we present next is simple. Direct routes must hit the target without falling into a comprehensive set. This is hard, hence these routes have to be quick - and the quickest way is to make least resistance steps. This will be made precise in the following.

First, since we have not assumed that P_ϵ is ergodic - to avoid triviality, we assume that

[6] Note that although in the examples $x \notin B$ we do allow $x \in B$. This enables us to analyze paths that leave x and subsequently return to x which is useful in proving subsequent results about interim dynamics.

$\mathcal{A}(x,B,W) \neq \emptyset$. An important observation is that there are typically many direct routes. Specifically, if there exists some $z, z' \notin W \cup \{x\}$ with $r(z,z') < \infty$ then $\mathcal{A}(x,B,W)$ is countably infinite. An important fact proven in Appendix 1 is that if $A \subseteq \mathcal{A}(x,B,W)$ then $r(A) = \min_{a \in A} r(a)$ is well-defined (and finite) - it is the least resistance of any path in the set A. We also define $t(A) = \min\{t(a) | a \in A, r(a) = r(A)\}$ to be the minimum number of transitions of any least resistance path in the set A.

The main result on direct paths characterizes their probability and length in terms of $r(A), t(A)$, for $A \subseteq \mathcal{A}(x,B,W)$. Proof is in Appendix 1.

Theorem 1. *There are constants $D_1(r), D_2(k,r)$ with $C \leq D_1(r), D_2(k,r) < \infty$ such that $C^{t(A)}\epsilon^{r(A)} \leq P_\epsilon(A|x) \leq D_1(r(A))\epsilon^{r(A)}$; and $E[t^k(a)|x,A] \leq D_2(k, r(A))/C^{t(A)}$.*

In particular, positive resistance direct routes are not very likely to occur as ϵ gets small, yet they are unlikely to be terribly long in the sense that the moments of the length are bounded independent of ϵ. Intuitively, at each point along a direct route there is a zero resistance path that leads to the forbidden set W. The more time spent along the route, the greater the danger that the path will fall into the forbidden set and fail to reach its destination. By contrast, we will see subsequently that the expected time spend in an irreducible class goes to infinity as $\epsilon \to 0$.

The probability bounds in Theorem 1 directly imply the two other facts characterizing direct routes.

Corollary 1. *Let $A = \{a | r(a) = r(\mathcal{A}(x,B,W))\}$ denote the least resistance paths in $\mathcal{A}(x,B,W) \neq \emptyset$. Then $\lim_{\epsilon \to 0} \frac{P_\epsilon(A|x)}{P_\epsilon(\mathcal{A}(x,B,W) \backslash A|x)} = \infty$.*

In other words, least resistance direct paths are far more likely than other direct paths. Applying this to the illustrative example yields facts (1) and (2) concerning transitions between the ergodic sets $\{0\}$ and $\{5\}$: by Theorem 1 these transitions are likely to be short, and typically of least resistance. Moving from $\{0\}$ to $\{5\}$ a least resistance path must hit $\{1\}$ and while it may remain at $\{1\}$ it must not return to $\{0\}$ before hitting $\{2\}$ - because going from $\{0\}$ to $\{1\}$ would add resistance to the transition. For the same reason, once $\{2\}$ is hit the path may not return to $\{1\}$. Similarly in moving from $\{5\}$ to $\{0\}$ least resistance paths must pass through $\{4\}$ then $\{3\}$ then $\{2\}$ in that order, and not return to $\{4\}$ from $\{3\}$ nor to$\{3\}$ from $\{2\}$.

As the next corollary shows, it is also the case that all least resistance direct paths have a probability similar to each other.

Corollary 2. *Let $A = \{a | r(a) = r(\mathcal{A}(x,B,W))\}$ and $a \in A$. Then $\frac{P_\epsilon(a|x)}{P_\epsilon(A|x)} \geq C^{t(a)}/D_1(r(A))$.*

This completes the discussion of the basic results on direct routes.

4. Transitions Between Irreducible Classes

We start again with an initial point $x \in Z$ with $\Omega(x) \neq \emptyset$, a forbidden set $W \subseteq Z$ and a target set $B \subseteq W$. Notice that the definition in section 3 implies that direct routes from x to B with forbidden set W are not allowed to pass through all points in $\Omega(x)$, since the forbidden set W was assumed to be comprehensive. We now wish to relax that restriction, and consider routes which are allowed to linger freely inside $\Omega(x)$.[7] So we exclude $\Omega(x)$ from the forbidden set, that is we assume $W \cap \Omega(x) = \emptyset$. Thus W cannot be comprehensive. However, we assume that W contains at least one point from every irreducible class except for $\Omega(x)$. We then call W *quasi-comprehensive*, and the paths $\boldsymbol{A}(x, B, W)$ from x to B with forbidden set W which have positive probability for $\epsilon > 0$ *quasi-direct routes.*

Ellison (2000) observes that being able to pass through every point in an irreducible class may have a profound impact on the nature of the paths. The main result of this section makes this precise by showing that before leaving $\Omega(x)$ for good, quasi-direct routes spend most of the time *within* $\Omega(x)$.

As in the direct case we assume the set $\boldsymbol{A}(x, B, W)$ is non-empty. Again we are interested in the structure of the paths in A, in particular: which paths in $\boldsymbol{A}(x, B, W)$ are most likely, what do these paths look like, and how long are they?

Let a be a path in $\boldsymbol{A}(x, B, W)$. It is convenient to view such a path as consisting of two distinct parts, the initial wandering in or near $\Omega(x)$ and the final crossing to B. We can think of this in terms of returning to x a number of times before leaving x to hit B without returning. In particular, let $\boldsymbol{A}^-$ denote the set of paths that begin and end at x and do not touch W, and let $\boldsymbol{A}^+$ be the routes from x to B that do not touch W nor x in between - that is the direct routes to B with forbidden set $W \cup \{x\}$. Notice that since W is quasi-comprehensive $W \cup \{x\}$ is comprehensive, so these are indeed direct routes. Then we have the unique decomposition of a into a^-, a^+ with $a^- \in \boldsymbol{A}^-$ and $a^+ \in \boldsymbol{A}^+$. We refer to a^- as the *equilibrium path* and a^+ as the *exit path.*

The equilibrium paths $\boldsymbol{A}^-$ have an additional very useful structure: a path $a \in \boldsymbol{A}^-$ can be decomposed into loops that begin and end at x but do not hit x (besides not hitting W) in between. That is, let $\mathcal{A}^0 = \mathcal{A}(x, \{x\}, W \cup \{x\})$ be the direct paths from x to $B = \{x\}$ avoiding the comprehensive set consisting of the quasi-comprehensive set W plus $\{x\}$ itself (the superscript 0 suggesting a loop). Then paths in $\boldsymbol{A}^-$are exactly sequences $a_1, a_2, \ldots, a_n$ such that $a_i \in \mathcal{A}^0$. We write $n(a)$ for the number of loops of a. (Note that it may be that $n(a) = 0$.)

So any path $a \in \boldsymbol{A}(x, B, W)$ has a unique decomposition $a_1, a_2, \cdots, a_n, a^+$ where the $a_i \in \mathcal{A}^0$ are the loops in $\mathcal{A}^0$ and $a^+ \in \boldsymbol{A}^+$ is the exit path to B. For non-trivial paths we then define the *equilibrium resistance* $\rho^-(a) = \max r(a_i)$, the *exit resistance* $\rho^+(a) = r(a^+)$ and the *peak resistance* $\rho(a) = \max\{\rho^-(a), \rho^+(a)\}$. For any $A \subseteq \boldsymbol{A}(x, B, W)$ we can define the *least peak resistance* $\rho(A) = \min_{a \in A} \rho(a)$. The first thing to understand is that least peak resistance paths are also least exit resistance paths:

Theorem 2. $\{a | \rho(a) = \rho(\boldsymbol{A}(x, B, W))\} \subseteq \{a | \rho^+(a) = \rho(\boldsymbol{A}(x, B, W))\}$

Proof. Since by definition $\rho(a) \geq \rho^+(a)$ the lemma can fail only if there is a path for which $r(a^+) < \rho(\boldsymbol{A}(x, B, W))$. But $x \in \Omega(x)$ by assumption, so there is a $y \in \Omega(x)$ (possibly $y = x$)

[7] Notice that for the case of singleton $\Omega(x)$ this means the path may remain at x for some time, or leave and return a number of times before hitting the target.

with $r(x,y)=0$. Hence the path $(x,y,a^+) \in \boldsymbol{A}(x,B,W)$ has peak resistance $\rho(x,y,a^+) = r(a^+) < \rho(\boldsymbol{A}(x,B,W))$ contradicting the fact that $\rho(\boldsymbol{A}(x,B,W))$ was the least peak resistance. $\square$

Hence a least peak resistance path - those that as we will show are the ones we are likely to see - consists of two parts: a^- which we will study later, and a^+which is a least resistance direct route from x to B with forbidden set $W \cup \{x\}$. Notice that some points $y \in \Omega(x)$ may support lower resistance direct routes to B with forbidden set $W \cup \Omega(x)$, that is, they are more likely to get there without returning to $\Omega(x)$. We call these *express exits.* Then a least resistance path from x to B must leave $\Omega(x)$ through an express exit: since the express exit is also in $\Omega(x)$ it can be reached from x with zero resistance. Hence to leave $\Omega(x)$ through any other exit would incur higher resistance.

The following result (proof in Appendix 2) plays a role in the theory of quasi-direct routes similar to that played by least resistance in the theory of direct routes in Corollary 1:

Theorem 3. *Let $A = \{a | \rho(a) = \rho(\boldsymbol{A}(x,B,W))\}$ denote the least peak resistance paths in $\boldsymbol{A}(x,B,W) \neq \emptyset$. Then $\lim_{\epsilon \to 0} \frac{P_\epsilon(A|x)}{P_\epsilon(\boldsymbol{A}(x,B,W) \backslash A|x)} = \infty$.*

Theorem 3 not only tells us the most likely route from $\Omega(x)$ to $\Omega \backslash \Omega(x)$, by implication it tells us where we are likely to end up in $\Omega \backslash \Omega(x)$. Let Ω_ρ be the irreducible classes in $\Omega \backslash \Omega(x)$ that are directly reachable from some point $y \in \Omega(x)$ with least resistance ρ and let $\Omega_{-\rho} = \Omega \backslash (\Omega_\rho \cup \Omega(x))$. Let $P_\epsilon(\Omega_i|x)$ denote the probability that starting at x the first arrival at $\Omega \backslash \Omega(x)$ is in Ω_i for $i = \rho, -\rho$. Then we have the following immediate corollary of Theorem 3.

Corollary 3. $\lim_{\epsilon \to 0} \frac{P_\epsilon(\Omega_\rho|x)}{P_\epsilon(\Omega_{-\rho}|x)} = \infty$.

We now study the set $\boldsymbol{A}^-$ of equilibrium paths. We know from Theorem 1 that transition paths in the direct route case are short. For paths that are allowed to remain in $\Omega(x)$ we have the opposite result: these paths are quite long. For the assertion to make sense it must be the case that we reach B with probability one, that is $P_\epsilon(\boldsymbol{A}(x,B,W)|x) = 1$. We assure this by assuming that $B = W$.[8] Our goal is to show that: $\boldsymbol{A}^-$ has paths of expected length $\epsilon^{-r(A^+)}$; that the fraction of time spent in $\Omega(x)$ goes to one; and that the absolute time spent outside of $\Omega(x)$ goes to infinity.

Recall that $a^- \in \boldsymbol{A}^-$ is a sequence $a_1, a_2, \ldots, a_n$ with $a_i \in \mathcal{A}^0$ loops at x. Now let $M(A^0)$ be the number of loops that lie in $A^0 \subseteq \mathcal{A}^0$. That is, if $f : \mathcal{A}^0 \to \Re$ is the indicator of A^0 ($f(a^0) = 1$ if $a^0 \in A^0$ and $f(a^0) = 0$ for $a^0 \in \mathcal{A}^0 \setminus A^0$) then $M(A^0)$ is the aggregate f^-. Also, let t_- be the amount of time along a^- spent outside of $\Omega(x)$. In Appendix 2 it is shown that the following holds:

Theorem 4. *If $B = W$ we have for some $C_1, D_3 > 0$*

$$1 + C_1 \epsilon^{-r(\boldsymbol{A}^+)} \leq E[t|x, \boldsymbol{A}(x,B,W)] \leq D_2(1,0)C^{-N} + D_3 C^{-2N} \epsilon^{-r(\boldsymbol{A}^+)}$$

and for $A^0 \subseteq \mathcal{A}^0$

$$C_1 C^{t(A^0)} \epsilon^{r(A^0) - r(\boldsymbol{A}^+)} \leq E[M(A^0)|x, \boldsymbol{A}^-] \leq D_1(r(A^0)) C^{-2N} \epsilon^{r(A^0) - r(\boldsymbol{A}^+)}.$$

[8] Recall that the paths in $\boldsymbol{A}(x,B,W)$ by definition have positive probability of reaching B without touching W along the way; when $B = W$ there is no other way of reaching B so this probability becomes one.

Moreover, $\lim_{\epsilon \to 0} E[t_-/t|x, \boldsymbol{A}^-] = 0$.

The result says that quasi-direct paths are long and that the paths return to x many times, and that during these excursions most of the time is spent within $\Omega(x)$. Moreover if there is some $a^0 \in \mathcal{A}^0$ with $r(a^0) < r(\boldsymbol{A}^+)$ then the amount of time spent outside of $\Omega(x)$ is large in an absolute sense. When applied to the coordination game of Section 2 the result yields assertions 3 and 4 made there, that going from $\{0\}$ to $\{5\}$, before $\{5\}$ is reached the system will spend most of the time at $\{0\}$ but will many times reach state $\{1\}$ for relatively brief periods; and that symmetrically, going from $\{5\}$ to $\{0\}$ states $\{4\}, \{3\}$ are visited for brief periods while most of the time the system is in state $\{5\}$.

5. The Fall of the Qing Dynasty

We now discuss how our results can be tested and used to interpret historical facts concerning sequences of long-run social events of small probability. This is the natural field of application of our theory, which concerns transitions along paths whose steps are each quite unlikely to occur. We focus in particular on the fall of the Qing dynasty in nineteen century China, using a variation of the model of Levine and Modica (2013).

There are L units of land and a finite number J of societies. In each period t each society j has one of a finite number of internal states $\xi_{jt} \in \Xi_j$. These states evolve according to a fixed Markov process $Q_j(\xi_{jt}|\xi_{jt-1}) > 0$ independent of ϵ. External forces such as disease, climate, other real shocks to productivity, or the interference of outsiders who are are protected themselves by geographical barriers, or superior technology can lead to changes in the internal state; the state may also represent changes to the internal structure of institutions. A good example of the Q_j process within a given unit of land can be found in Acemoglu and Robinson (2001): there external shocks (recessions) drive changes in institutions - the franchise is extended or contracted. The ability of a society to resist and influence other societies is indexed by state power $\gamma_j(\xi_{jt})$. Societies may or may not satisfy incentive constraints: we represent this by a stability index $b_j \in \{0, 1\}$ with 1 indicating stability, where societies violating incentives are thought to be unstable. This simplifies Levine and Modica (2013) by making the learning dynamics part of the evolutionary process. A state z is a list of land holding and real shocks of the different societies, $z = (L_1, \xi_1, L_2, \xi_2, \ldots, L_J, \xi_J)$.

We now describe P_ϵ. We assume that at most one unit of land changes hands each period. The probability that society j loses a unit of land is given by a conflict resolution function determined by the probabilities $\pi(b_j, \gamma_j, L_{jt}, b_{-j}, \gamma_{-j}, L_{-jt})[\epsilon]$ of j losing one unit of land.[9] Notice that since at most one unit of land can change hands each period $\sum_{j=1}^{J} \pi_{jt} \leq 1$ and the shocks must necessarily be correlated. If a unit of land is lost it is gained by a society chosen randomly according to the function $\lambda(k|j, L_t) > 0$ for $k \neq j$ and $\lambda(j|j, L_t) = 0$.

[9]Notice that this slightly generalizes Levine and Modica (2013) by allowing the probability to depend on the stability status of other societies. This makes it easier to deal with the adding up constraint that only one unit of land can be lost each period.

We make several assumptions about the conflict resolution function beyond regularity. We assume that π is strictly positive if $\epsilon > 0$ and that it is symmetric, so that the names of the societies do not matter, only their land holdings, state power and stability. We assume that it is monotone, so that the probability of j losing land decreases with its own state power and land, and increases with that of other societies. We assume that unstable societies always have resistance zero to losing a unit of land - if incentive constraints are not satisfied, individuals experiment with different actions, and societies experiment with different institutions (albeit just on a single unit of land at a time). For stable societies we assume the resistance is $r(\gamma_j, L_j, \gamma_{-j}, L_{-j})$ which is strictly monotone when non-zero, and that the weakest society society with positive land holding has zero resistance to losing a unit of land. Finally we assume for given land holding that resistance is greater when facing more than one opponent with positive land holding than when all enemy land is in the hands of the strongest land holding opponent - also that this is strict if resistance is positive. We assume that there is some stable society strong enough that resistance to losing land is positive when it holds all the land and that the strongest unstable society with the most favorable value of ξ_j is stronger than the strongest stable society in its most favorable value of ξ_j - reflecting the idea that unstable societies face weaker incentive constraints.

Call societies with positive land holding *active*. Since the weakest active society by assumption has zero resistance to losing land, and since losing land by monotonicity lowers its resistance to losing more land, we see that the only possible candidates irreducible sets are *hegemonies* - single societies that control all the land. Specifically if $y \in \Omega(x)$ then there is a single stable society $j(x)$ that controls all the land, and the different states $y \in \Omega(x)$ correspond to different shocks ξ. Depending on the resistance function r there may or may not be irreducible classes: so that the model is of interest, we assume that there are at least two.

Our interest here is in how hegemonies fall - that is, how we move from a hegemony $\Omega(x)$ to $B = \Omega \backslash \Omega(x), W = B$. [10]Accordingly, we may apply Theorem 3 to conclude that when ϵ is small it is nearly certain that this transition will take place along a least peak resistance path, and we want to describe such paths in the model at hand. In doing so the interesting exercise is to compare the theoretical predictions of the transition to the fall of an actual hegemony. As a case study for which there is quite a bit of historical information, we take the fall of the final Qing dynasty in China and the subsequent rise of the communist hegemony.[11]

The exit path from $\Omega(x)$ must be a direct path from $y \in \Omega(x)$ to B. To have least exit resistance, y must correspond to the value of $\xi_{j(x)}$ for which state power is the least, that is the worst possible shock. The direct path itself must have three stages: it must first move through the basin of $\Omega(x)$

[10]Another blocking set of interest consists of the states where the hegemony loses a certain threshold amount of land to an invader. The paths to this blocking set are qualitatively the same as those we are going to describe below so we will not elaborate on it further. However, such sets may be empirically important, because it is easier to say that a hegemony lost 1/3rd of its land than to say that it "left the basin."

[11]There are of course many accounts of this period, and while they sometimes disagree on exactly who did what to whom when, all agree on the basic facts we describe below. One readable account by a journalist is that of Fenby (2008).

- we call this the *fall of the hegemony.* After the fall of the hegemony there should be a turbulent period of *warlordism* in which the basin of a different hegemony has not yet been reached. Then there will be a *consolidation* in which the basin of the eventual new hegemony is crossed until the new hegemony is reached.

During the fall of the hegemony unstable societies with maximum state power - following Levine and Modica (2013) we call them *zealots* - play a key role. The reason for this is simple: while in the basin of $\Omega(x)$ the hegemony has a positive resistance to losing land. By monotonicity it has the least resistance to losing land if all the land not held by the hegemon is held by a single society with the greatest possible state power in its most favorable state - that is, zealots. Hence least resistance implies that land lost by the hegemony during the fall must be lost to zealots, and that once lost it is never regained.

Prior to the fall of the hegemony, by Theorem 4 the theory predicts that there should be a small fraction but large number of periods where there are failed rebellions: lands that are lost to other societies but quickly regained. These failed rebellions may or may not involve zealots, and need not take place when ξ_j is at its nadir. However, prior to the actual exit, ξ_j must be such that state power is at a nadir, and this means that resistance to rebellions is less, so there should be more frequent and larger rebellions prior to the final fall of the society.

To put this in the Chinese context, the basic fact is that Chinese institutions that lasted from roughly the introduction of the Imperial Examination System in 605 CE until 1911 CE were swept away in less than a year. It is useful to begin the story about 1838, before the First Opium War. At that time the Qing dynasty held a hegemony over China proper: the area bordered by the difficult terrain of Indochina in the Southeast, the Himalayan mountains in the South, the inhospitable deserts in the West, the Pacific Ocean in the East and the wasteland of Mongolia in the North. It also held a number of outlying areas not part of China proper - the Korean Peninsula, Indochina and Taiwan. As these are not so easily defended, are not Chinese, and have only been part of the Chinese hegemony at certain times - and moreover, the current government claims only Taiwan among these territories - we do not count them as part of the hegemony.

Several independent sources of instability concurred to the fall of the hegemony. In the early 1800s China fell into a severe economic depression from which it did not recover prior to the fall of the hegemony. Outsiders, most notably the English, French, and Japanese actively intervened in China, sometimes fighting for and other times against the Qing, but in any case certainly putting pressure on institutional change. Opium consumption, induced by the English to correct trade imbalances, increased to an unpredictable, and dramatic level.

From 1839 to 1910 there were a series of unsuccessful attempts to overthrow the Qing dynasty including local rebellions and acts of defiance by committed revolutionaries. During this time the outlying territories were lost: Korea became independent, Indochina was lost to the French, and Taiwan to the Japanese further weakening the hegemon. Roughly speaking the state ξ_j became increasingly worse. However, each internal rebellion was successfully repressed, each war brought to an end, and in each case the Qing hegemony over China proper - tax collecting authority, control of institutions local and global - remained intact. There were institutional changes that took place during this period, some forced by the outsiders, and an attempt to placate the revolutionaries, such as the abolition of the imperial examination system in 1905. These can be viewed as shocks ξ_j that further weakened the state. Although it is hard to measure the relative frequency of failed rebellions before and after the economic weakness of the 19th Century there seem not to have been such dramatic episodes as the Duggan revolt which lasted for several years. As Theorem 4 predicts, before the actual fall the state ξ_j is very bad, and there are many and probably increasing failed attempts at rebellion.

The actual fall of the Qing occurred in 1911 and as Theorem 1 suggests, it was very quick. There were again a series of revolts - now however they succeed. They are loosely coordinated by Sun Yat Sen. The groups carrying out these revolts can reasonably be described as zealots: they share in common a dedication to overthrowing the Qing, they are willing to suffer severe risk and live under unpleasant circumstances in order to achieve that goal. Such behavior is power maximizing - but is not stable in the sense that no society has every lasted very long based on the fanatical devotion of its members - nor was it the case in China. Hence the theoretical description of the fall of the hegemony is relatively accurate: zealots quickly capture the land, and do so without a serious setback. By the end of 1911 the Qing Emperor abdicated and Sun Yat Sen became the provisional President of China, which however no longer was hegemonic in any reasonable sense of the word.

Next is the period of warlordism, both in theory and in fact. In theory, once out of the basin, least resistance paths all must have resistance zero. Hence very little can be said in general: there can be many competing societies, land may be lost and gained, zealots may or may not play a role. The only general proposition is that no stable society has yet gained enough land to reach the basin of its own hegemony. Again, this is an accurate description of the situation in China between 1911 and 1946. Sun Yat Sen was quickly deposed by a less fanatical and more materialistic warlord Yuan Shikai, but until about 1927, and even after, there are many warlords in various parts of China who rise and fall, many revolutions, some successful and other unsuccessful. There is also the Sino-Tibetan war and the Soviet invasion of Xinjiang during this period. Basically the theory predicts chaos (in the non-technical sense) and that is what we see. Beginning in about 1927 things settle down slightly with two relatively more powerful groups, the Nationalists and the Communists, fighting a civil war - but there remain many warlords who continue to rise and fall, at times forming alliances or professing allegiance to the two more significant groups. These two groups, unlike the earlier revolutionaries appear to have coherent and potentially stable institutions. Then in 1936 the Japanese seize control of most of the country, an occupation that lasts until 1945.

The final stage of a least resistance transition is the consolidation. Again all transitions must have zero resistance, but now we are in the basin of the hegemony so the least resistance path consists of the hegemony gaining territory - without losing any - until hegemony is again established. Notice that since in this model once the basin is left there are zero resistance transitions to any particular hegemony breaching the threshold, the model makes no prediction about which hegemony eventually emerges - in particular there is a non-negligible probability that even a very weak hegemony emerges. In China, the threshold appears to be reached about 1946 when the Communists controlled about a quarter of the country and about a third of the population. They quickly overran the remaining areas held by the Nationalists, who retreated to Taiwan in 1949.

6. The Big Picture

Consider an overview of the dynamics of P_ϵ. Starting at any point x by Theorem 1 we move quickly to one of the irreducible sets $\Omega(y)$. Once there, by Theorem 4 there it is a long time before we reach a different $\Omega(z)$ and most of that time is spent in $\Omega(y)$. One question we wish to address is what the dynamics look like during the long period when we are in $\Omega(y)$. When we do finally leave $\Omega(y)$ by Proposition 2 we move quickly to the next $\Omega(z)$ and it is most likely the irreducible set that has least exit resistance from $\Omega(y)$. The second question we wish to address is over the longer run how much time do we spend in the different irreducible sets $\Omega(y)$. To this end, we assume in this section that P_ϵ is ergodic for $\epsilon > 0$ and denote by μ_ϵ the unique ergodic distribution of the process.

6.1. Transitory and Irreducible Dynamics

When $\epsilon = 0$ we cannot move between irreducible sets $\Omega(x)$ but we have well defined and ergodic dynamics within each such set. Moreover, these dynamics are fast in the sense that if we are interested in the approximate probability of events, it suffices to consider paths of bounded length (where the bound is obviously independent of ϵ which plays no role in the dynamics of P_0). For such events, the probabilities in P_0 are much the same as for P_ϵ for ϵ small. Specifically,

Theorem 5. *let A_1 and A_2 be any collections of paths of bounded length and for which $P_0(A_2|x) > 0$. Then*

$$\lim_{\epsilon\to 0} \frac{P_\epsilon(A_1|x)}{P_\epsilon(A_2|x)} = \frac{P_0(A_1|x)}{P_0(A_2|x)}.$$

Proof. Since the probabilities are defined by finite sums of finite products of the transition probabilities $P_\epsilon(z|y)$ and the length of the sums and products are bounded independent of ϵ the result follows immediately from the assumption that $\lim_{\epsilon\to 0} P_\epsilon(z|y) = P_0(z|y)$. □

In particular since paths that lie entirely within $\Omega(x)$ have probability one in P_0 given x the probability of sets of paths of finite length within $\Omega(x)$ is roughly the same in P_ϵ as in P_0 when ϵ is small. For example, in the Acemoglu and Robinson (2001) example, the pattern and probability of recessions and government transitions is roughly the same for all small or zero ϵ.

The other important characteristic of $\Omega(x)$ is the amount of time spent at different points - in the Acemoglu and Robinson (2001) case we might be interested in the frequency of recessions, for example. Notice that if we restrict the state space to $\Omega(x)$ then P_0 is an ergodic Markov process on that space, so has a unique and strictly positive ergodic distribution $\overline{\mu}_0(y)$, where $\sum_{y\in\Omega(x)} \overline{\mu}_0(y) = 1$. Notice in particular that if $y \in \Omega(x)$ the ratio $\overline{\mu}_0(x)/\overline{\mu}_0(y)$ is well-defined and finite. We can relate this to the ratio of stationary probabilities $\mu_\epsilon(x)/\mu_\epsilon(y)$ for the process when $\epsilon > 0$; in Appendix 3 we show:

Theorem 6. *If $y \in \Omega(x)$ then*

$$\lim_{\epsilon \to 0} \frac{\mu_\epsilon(x)}{\mu_\epsilon(y)} = \frac{\overline{\mu}_0(x)}{\overline{\mu}_0(y)}.$$

For any two irreducible classes $\Omega(x), \Omega(y)$ define the resistance $r(\Omega(x), \Omega(y))$ as the least resistance of any direct path from x to the target $\Omega(y)$ with forbidden set $W = \{x\} \cup \Omega \setminus \Omega(x)$ - paths which can stay in $\Omega(x)$ as long as they do not go back to x, and when they leave it they do not touch other irreducible sets. Note that $r(\Omega(x), \Omega(y))$ is independent of the particular starting point x in $\Omega(x)$ since there are zero resistance paths from any point in $\Omega(x)$ to any other. Let $\mathscr{W}$ be all the irreducible classes. Define also $r(\Omega(x)) = \min_{\Omega(y) \in \mathscr{W}} r(\Omega(x), \Omega(y))$. It is shown in Appendix 3 that:

Theorem 7. *Allowing that $\Omega(x)$ may be empty, if $A = \mathcal{A}(x, \{y\}, \{x\} \cup \{y\} \cup (\Omega \backslash \Omega(x))$ are the direct routes from x to y with forbidden set $\{x\} \cup \{y\} \cup (\Omega \backslash \Omega(x))$ then $\mu_\epsilon(y) \geq \mu_\epsilon(x) C^N \epsilon^{r(A)}$. There is also a constant D_4 such that if $x \in \Omega(x)$ and there is a zero resistance path from y to x then also $\mu_\epsilon(y) \leq \mu_\epsilon(x) D_4 \epsilon^{\min\{r(A), r(\Omega(x))\}}$.*

It is convenient to define the *inner basin* of $\Omega(x)$ as the set of points y that have zero resistance of reaching $\Omega(x)$ and in addition have resistance less than or equal to $r(\Omega(x))$ of being reached from x along a direct route with forbidden set $\{x\} \cup \{y\} \cup (\Omega \backslash \Omega(x))$. These are points that are in a sense "close" to $\Omega(x)$. For the inner basin, the bound in Theorem 7 provides an exact computation of the resistance, that is for y in the inner basin of $\Omega(x)$ and A the direct routes from x to y we have $\mu_\epsilon(x) C^N \epsilon^{r(A)} \leq \mu_\epsilon(y) \leq \mu_\epsilon(x) D_4 \epsilon^{r(A)}$. For points that are not in the inner basin of any irreducible set, that is, points that are in the outer basin of one or more irreducible set, but are "hard" to reach from any of them we cannot exact computation of the resistance of the stationary probabilities - however, Theorem 7 does show that such points are not terribly likely compared to points in the inner basin - and perhaps they are not of such great interest.

6.2. Long Run Ergodic Probabilities

Consider again our overview of the dynamics of P_ϵ. Starting at any point x by Theorem 1 we move quickly to one of the irreducible sets $\Omega(y)$. Once, by Theorem 4 there it is a long time before we reach a different $\Omega(z)$ and most of that time is spent in $\Omega(y)$. When we do finally leave $\Omega(y)$ by Proposition 2 we move quickly - and directly, in our sense - to the next $\Omega(z)$ and it is most likely the irreducible set that has least exit resistance from $\Omega(y)$. Proceeding in this way we get a sequence of irreducible sets $\Omega(x_i)$ connected by least exit resistances. Since the set $\mathscr{W}$ of irreducible classes in P_0 is finite, eventually this sequence must have a loop.

More general than the notion of a loop, we introduce the notion of a *circuit* on a subset $\Phi \subseteq \Psi$ of a set of points on which a resistance function $r(\psi, \phi)$ is defined. In the case above, where $\Psi = \mathscr{W}$, the relevant resistances are the least exit resistances connecting the irreducible class ψ with ϕ. For any $\psi \in \Psi$ we define the *least resistance* $r(\psi) = \min_{\phi \in \Psi} r(\psi, \phi)$. We say that Φ is a *circuit* if for each pair $\phi_1, \psi \in \Phi$ there is a path $\phi_1, \phi_2, \ldots, \phi_n \in \Phi$ with $\phi_n = \psi$ such that for $\tau = 2, 3, \ldots n$ we have $r(\phi_{\tau-1}, \phi_\tau) = r(\phi_{\tau-1})$, that is, there is a path from $\phi_1$to ψ in Φ such that each connection has least resistance.

Our basic observation is that once we reach a circuit, we remain within the circuit for a long time before going to another circuit. Given that we stay in ψ roughly $\epsilon^{-r(\psi)}$ periods before moving to another irreducible class in the circuit, we might expect that the amount of time we spend at ψ is roughly $\epsilon^{r(\phi)-r(\psi)}$ as long as the amount of time we spend at ϕ. In Appendix 4 we show that this is indeed true.

Theorem 8. *If the irreducible classes $\Omega(x)$ and $\Omega(y)$ are in the same circuit then*

$$\frac{C^N}{N^{N-1}D^N}\epsilon^{r(\Omega(y))-r(\Omega(x))} \leq \frac{\mu_\epsilon(x)}{\mu_\epsilon(y)} \leq \frac{N^{N-1}D^N}{C^N}\epsilon^{r(\Omega(y))-r(\Omega(x))}.$$

How long do we actually spend in a circuit? Since the probability of leaving ψ is of order $\epsilon^{r(\psi)}$ the expected length of any visit to ψ is $1/\epsilon^{r(\psi)}$. On the other hand the probability of going to a fixed ϕ out of the circuit is of order $\epsilon^{r(\psi,\phi)}$. Hence the probability of going to ϕ during a visit to ψ is of order $(1/\epsilon^{r(\psi)})\epsilon^{r(\psi,\phi)}$. In order for this to occur with very high probability the number of visits to ψ must be roughly k_ψ where $k_\psi(1/\epsilon^{r(\psi)})\epsilon^{r(\psi,\phi)} = 1$. That is $k_\psi = 1/\epsilon^{r(\psi,\phi)-r(\psi)}$. If we define the *modified resistance from ψ to ϕ* as $R(\psi,\phi) = r(\psi,\phi) - r(\psi)$, then the number of visits is least for the ψ which has minimum $R(\psi,\phi)$ over $\psi \in \Phi$. This is the most likely (actually least modified resistant) exit from the circuit. Also, it will exit to a circuit which is easiest to reach. This in turn suggests that we can form circuits of circuits using modified resistances as the measure of resistance in going from one circuit to another. The system moves between circuits of circuits in a longer time horizon. Moreover, as we have seen the crossings between circuits are direct routes, hence we will define resistance in terms of such paths.

Specifically we recursively define a class of reverse filtrations with resistances over the set $\Psi^0 = \mathscr{W}$ of irreducible sets for P_0; assume $\mathscr{W}$ has $N_{\mathscr{W}}$ elements, with $N_{\mathscr{W}} \geq 2$. As before, for $\psi, \phi \in \mathscr{W}$ the resistance $r^0(\psi,\phi)$ is the least resistance of any direct path from $x \in \psi$ to the target ϕ with forbidden set $W = \{x\} \cup \Omega \setminus \Omega(x)$. Now starting with Ψ^{k-1} we observe from Appendix 4 that there is at least one non-trivial circuit, and that every singleton element is trivially a circuit. Hence we can form a non-trivial partition of Ψ^{k-1} into circuits, and denote this partition Ψ^k. As before we define the modified resistance $R^{k-1}(\psi^{k-1},\phi^{k-1}) = r^{k-1}(\psi^{k-1},\phi^{k-1}) - r^{k-1}(\psi^{k-1})$, and the resistance function on Ψ^k by the least modified resistance: $r^k(\psi^k,\phi^k) = \min_{\psi^{k-1}\in\psi^k,\phi^{k-1}\in\phi^k} R^{k-1}(\psi^{k-1},\phi^{k-1})$. Note that since each partition is non-trivial, this construction has at most $k \leq N_{\mathscr{W}}$ layers before the partition has a single element and the construction stops.

Given a reverse filtration, for given $x \in \Omega(x)$ we can now define $\psi^k(x)$ recursively by $x \in \psi^o(x), \psi^0(x) \in \psi^1(x) \ldots$. The *modified radius* of $x \in \Omega(x)$ of order k is then defined by

$$\overline{R}^k(x) = \sum_{\kappa=0}^{k} r^\kappa(\psi^\kappa(x)).$$

Then we show in Appendix 4 that

Theorem 9. *Let k be such that $\psi^k(x) = \psi^k(y)$; then*

$$\frac{C^N}{N^{N-2}D^N}\epsilon^{\overline{R}^{k-1}(y)-\overline{R}^{k-1}(x)} \leq \frac{\mu_\epsilon(x)}{\mu_\epsilon(y)} \leq \frac{N^{N-2}D^N}{C^N}\epsilon^{\overline{R}^{k-1}(y)-\overline{R}^{k-1}(x)}.$$

It is useful to define $\overline{R}^{k-1}(y) - \overline{R}^{k-1}(x)$ as the *relative ergodic resistance* of x over y so that this theorem just says that the relative probabilities are proportional to ϵ to the power of the relative ergodic resistance. Notice that the states x that have ergodic probabilities that are bounded away from zero - the *stochastically stable* states - are exactly those with the least values of $\overline{R}^{k-1}(x)$ where k is the highest layer of the filtration at which the partition has a single element.

6.3. Examples

Turning first to the example of Section 2, we see that the two irreducible classes are necessarily on the same circuit, so by Theorem 9 the relative resistance is simply the difference in least resistances of the two points $\{0\}, \{5\}$, which are 2 and 3 respectively, so that the system spends roughly ϵ^{-1} more time at $\{5\}$ than at $\{0\}$. Note that since there are only two trees on a set with two points, this same result may easily be derived also from (for example) Young (1993). However, we may also apply Theorem 7 to see that $\{4\}$ also occurs about ϵ times as often as $\{5\}$, that is to say, roughly as frequently as $\{0\}$. From Theorem 4 we also see that while $\{0\}$ will be seen for long stretches of time, the state $\{4\}$ will be seen frequently but only briefly before reverting to $\{5\}$. This is the fifth and final assertion about the example made in Section 2.

To further illustrate the application of Theorem 9, let us give a complete analysis of the case where $\mathscr{W}$ has three elements. Note that there are 9 trees on 3 points, so the analysis by means of trees is already difficult. For simplicity let us make the generic assumption that no two resistances or sums or differences of resistances are equal.

There are two cases: either there is a single circuit, or there is one circuit consisting of two points, and a separate isolated point. The case of a single circuit is trivial - in this case the relative ergodic resistances are given simply by the differences in least resistances between the three points, and the stochastically stable state is the point with least least resistance.

Take finally, the case of $\mathscr{W}$ with one two-point circuit and an isolated point, and denote by ϕ_a, ϕ_b the two points on the circuit with ϕ_c the remaining point. Assume without loss of generality that $r(\phi_a) > r(\phi_b)$ so that within the circuit ϕ_a is relatively more likely. Notice that $r(\phi_a, \phi_b) < r(\phi_a, \phi_c), r(\phi_b, \phi_a) < r(\phi_b, \phi_c)$ since ϕ_a, ϕ_b are on the same circuit - this also implies $r(\phi_a) = r(\phi_a, \phi_b), r(\phi_b) = r(\phi_b, \phi_a)$. Turning to the recursion, we need to work out the least modified resistances. Let $\psi_a = \{\phi_a, \phi_b\}$ be the circuit and $\psi_c = \phi_c$ the isolated point. Then $r^1(\psi_c, \psi_a) = \min\{r(\phi_c, \phi_a) - r(\phi_c), r(\phi_c, \phi_b) - r(\phi_c)\} = 0$ while $r^1(\psi_a, \psi_c) = \min\{r(\phi_a, \phi_c) - r(\phi_a), r(\phi_b, \phi_c) - r(\phi_b)\}$. Hence $\overline{R}^1(\phi_c) = r(\phi_c)$, which is just the radius of ϕ_c, while

$$\begin{aligned}
\overline{R}^1(\phi_a) &= r(\phi_a) + \min\{r(\phi_a, \phi_c) - r(\phi_a), r(\phi_b, \phi_c) - r(\phi_b)\} \\
&= \min\{r(\phi_a, \phi_c), r(\phi_a) + r(\phi_b, \phi_c) - r(\phi_b)\}
\end{aligned}$$

which is to say exactly what Ellison (2000) defines as the modified co-radius of ϕ_c.[12] The relative ergodic resistance of ϕ_a over ϕ_c is therefore $r(\phi_c)-\overline{R}^1(\phi_a)$, while the relative ergodic resistance of ϕ_b can be recovered from the relative ergodic resistance of ϕ_b over ϕ_a which is just $r(\phi_a)-r(\phi_b)$. With respect to stochastic stability, we see that ϕ_c is stochastically stable if and only if its radius $r(\phi_c)$ is greater than its co-radius $\overline{R}^1(\phi_a)$ which is Ellison (2000)'s sufficient condition, and otherwise ϕ_a is stochastically stable. In short, the entire ergodic picture comes down to computing three numbers: the radius and co-radius of ϕ_c and the difference between the radii of ϕ_a and ϕ_b.

7. Conclusion

This paper is about events and combinations of events that are unlikely and that can be modeled as a finite Markov process, in particular how such a process moves from one relatively stable long-run state to another. Examples are transitions between different equilibria in a game or different political regimes. We show that these systems exhibit long periods of stability punctuated by brief episodes of change, and we give a detailed description of the probabilities and frequencies of these different outcomes. Within the literature on "evolution of conventions" we complement the results of Kandori, Mailath and Rob (1993), Young (1993) and Ellison (2000) on long run dynamics in games. When applied to the context of social evolution, our theory has implications both for the societies we are likely to see and for the design of institutions: institutions that will persist for long periods of time must be robust against multiple failures, and it is these multiple failures that lead a society to fall.

It may be useful to look at smaller systems about which we have a great deal of information - also subject to small unlikely shocks, and subject to the same type of Markov analysis - to see what is involved. For example commercial airlines, which despite the vast number of flights and miles flown crash relatively infrequently. As our theory predicts, when they do crash, it is typically due to multiple near simultaneous failures. To take a specific example, on November 24, 2001, en route from Berlin on approach to Zurich Crossair Flight 3597 crashed near Bassersdorf, Switzerland killing 24 of the 33 people on board. According to the flight investigation seven independent unfortunate events occurred on that occasion.[13] These multiple failures seem typical of commercial aviation crashes. Each individual failures is unlikely, but none terribly so. What is highly unlikely is that all occur in combination. In general airplanes are designed with a high degree of redundancy to provide insulation against failure of one or even several components: multiple pilots, multiple navigation

[12] In fact the modified co-radius is defined as the larger of $\overline{R}^1(\phi_a)$ and $\overline{R}^1(\phi_b) = \min\{r(\phi_b,\phi_c), r(\phi_b,\phi_a)+r(\phi_a,\phi_c)-r(\phi_a,\phi_b)\}$. However, $r(\phi_a,\phi_c) > r(\phi_b,\phi_a)+r(\phi_a,\phi_c)-r(\phi_a,\phi_b)$ and $r(\phi_a,\phi_b)+r(\phi_b,\phi_c)-r(\phi_b,\phi_a) > r(\phi_b,\phi_c)$ imply $\overline{R}^1(\phi_a) \geq \overline{R}^1(\phi_b)$.

[13] See AAIB (2002): (1) the pilot had a bad record of following procedures during landing and was inadequately trained, but was allowed never-the-less to transport passengers; (2) the flight was behind schedule and consequently the pilot was in a hurry to land; (3) due to noise regulations the plane was diverted to a less safe runway; (4) the runway had inadequate instrumentation and the airport parameters and protocols for landing on the runway were inadequate; (5) the range of hills the plane crashed into was not marked on the chart; (6) the pilot put the plane into an overly steep descent and descended too low without proper visual contact with the ground; (7) the pilot did not monitor the proper instruments during the attempted landing.

systems, multiple engines, multiple independent hydraulic systems and so forth. So it is with human societies. Those that survive for long periods of time must be well cushioned against even multiple failures. For example, the fall of the Roman Empire has been attributed to many factors: religious ferment, the plague, corruption, the forced migration of hostile outsiders, economic recession, and so forth. Despite the effort of historians to establish each as "the" cause of the fall, as is the case with Flight 3597 all of these things happened - and while each is uncommon, none is particularly unlikely, and the Roman Empire had suffered through each of these, often in combination, many times before. What is unique about the fall is that all these things occurred at once. When a system or society is well designed it takes a perfect storm - everything going wrong at once - to bring it down. But - as this paper shows - it is the least unlikely combination of things - the least resistance direct route - that will typically lead - for good or ill - to abrupt and sudden change.

Appendix 1: Direct Routes

Our goal is to establish probability and expectations bounds on subsets $A \subseteq \mathcal{A}(x, B, W) \neq \emptyset$, in particular to prove Theorem 1 in the text. It will be convenient to define $\overline{r}$ to be the largest finite value of $r(y, z)$ and $\underline{r}$ to be the smallest non-zero value of $r(y, z)$. Since the state space is finite these are well-defined. Once we establish that $r(A) = \min_{a \in A} r(a)$ exists, getting a lower bound on $P_\epsilon(A|x)$ is relatively easy: it is bounded below by the probability of a path $a \in A$ with resistance $r(a)$, which is to say, it of order $\epsilon^{r(a)} = \epsilon^{r(A)}$. The main goal is to establish a similar upper bound. The problem is that A can easily contain infinitely many paths with resistance $r(A)$ as well as paths of greater resistance. However, there are only finitely many paths of any given length, so if there are infinitely many paths most of them must be very long. The idea is that since paths in A must avoid the comprehensive set W they are not likely to be very long since there are zero resistance routes to W. To make this precise we construct a finite set of template paths of relatively low resistance and show that all the paths in A can be constructed by adding loops to the template paths. We then show that the probability of all paths constructed from a given template is bounded by the probability of the template times a constant that does not depend on ϵ.

Since an analysis of loops form a key part of the analysis, we begin by introducing the notion of loop-cutting. If $a = (z_0, z_2, \ldots, z_t)$ we say that a' is a *loop-cut* of a at $z_\tau = z_{\tau'}$ for $\tau' > \tau$ if $a' = (z_0, z_1, \ldots, z_\tau, z_{\tau'+1}, \ldots, z_t)$ for $\tau' < t$, and $a' = (z_0, z_1, \ldots, z_{\tau-1}, z_\tau)$ if $\tau' = t$. Note the obvious fact that $r(a') \leq r(a)$. A map m from the set of all paths to itself is a *loop-cutting algorithm* if there is a sequence $a_1, a_2, \ldots, a_M$ with $a_1 = a$, $a_M = m(a)$ and a_{j+1} is a loop-cut of a_j for $j = 1, 2, \ldots, M-1$. Note that $r(m(a)) \leq r(a)$. The idea is that $m(a)$ is a template for a from which a can be reconstructed by adding loops.

A loop cutting algorithm is *maximal* if $m(m(a)) = m(a)$. Recall that $t(a)$ denotes the number on transitions in a; we let $t(m)$ denote the greatest number of transitions of any path in $m(\mathcal{A}(x, B, W))$.

We can establish the existence of $\min_{a \in A} r(a)$ using the *zero-cut* algorithm. For any $a = (z_0, z_2, \ldots z_t)$ if there is no loop of zero resistance stop. Otherwise cut the first and shortest loop of zero resistance and repeat.[14] This is obviously maximal. Note that $m(a)$ is a *no-zero-loop* path in the sense that it contains no zero-resistance loops, and that $r(m(a)) = r(a)$. Our first step is to give a bound on the length of no-zero-loop paths.

Lemma 1. *If a is a no-zero-loop path then $t(a) \leq (N-1)r(a)/\underline{r}$.*

Proof. Observe that since non-zero resistance transitions have resistance at least $\underline{r}$ there are at most $r(a)/\underline{r}$ such transitions, and the remaining transitions must have zero resistance. Since there are no zero-resistance loops, the number of zero-resistance transitions between each positive resistance transition is at most $N-1$. □

We can now apply the zero-cut algorithm to prove the basic fact that

Lemma 2. *$r(A) = \min_{a \in A} r(a)$ is well-defined.*

Proof. Fix $a \in A$, noting that $r(a) < \infty$. Let m be the zero-cut algorithm. Consider that for any $a' \in A$ with $r(a') \leq r(a)$ we have $r(a') = r(m(a'))$ and that $t(m(a')) \leq (N-1)r(a)/\underline{r}$. But there are only finitely many paths of this length, so only finitely many possible values of $r(a') \leq r(a)$. Hence $\min_{a \in A} r(a)$ exists. □

We can now establish the "easy" lower bound on $P_\epsilon(A|x)$. Recall that $t(A) = \min\{t(a)|a \in A, r(a) = r(A)\}$.[15]

Lemma 3. *$P_\epsilon(A|x) \geq C^{t(A)}\epsilon^{r(A)}$.*

Proof. Let $a \in A$ satisfy $r(a) = r(A)$ and $t(a) = t(A)$.Then

$$P_\epsilon(A|x) \geq P_\epsilon(a|x) = \prod_{\tau=1}^{t(a)} P_\epsilon(z_\tau|z_{\tau-1}) \geq \prod_{\tau=1}^{t(a)} C\epsilon^{r(z_\tau, z_{\tau-1})} = C^{t(a)}\epsilon^{r(a)}.$$

□

To establish an upper bound, we start by establishing the fact that, given that W is comprehensive, long loops are not very likely. Let $L(z)$ be the set consisting of all finite resistance loops $(z, z_1, z_2, \ldots, z_{t-1}, z)$ such that $z_\tau \notin W$.

Lemma 4. $P_\epsilon(t(a)|L(z)) \leq \left[(1-C^N)^{1/N}\right]^{t(a)-N}$

Proof. By convention if $t = 0$ then $P_\epsilon(t|L(z)) \leq 1$. Otherwise, block t into $\lfloor t/N \rfloor$ blocks of length N ($\lfloor t/N \rfloor$ being the largest integer not greater than t/N). Since W is comprehensive there is a positive probability path from z to W when $\epsilon = 0$. It follows that there must be such a path with no more than N transitions. From the definition of resistance, such a path must have zero resistance, so that each transition has probability at least C regardless of ϵ. Hence the probability starting at z of hitting W in N steps is at least C^N. Since the paths in question do not in fact hit W we conclude that $P_\epsilon(t|L(z)) \leq (1-C^N)^{\lfloor t/N \rfloor} \leq (1-C^N)^{((t/N)-1)}$. □

[14] In other words, if there are multiple loops at a point we only cut the first.

[15] This minimum exists because the set is a set of non-negative integers.

We now want to reverse the loop-cutting procedure by adding loops to templates to construct the paths in A. To do so for $a \in m(A)$ we will define a map $\overline{m^{-1}}(a)$. Observe that since m is a loop-cutting algorithm, if $a = (z_0, z_1, \ldots, z_t) \in m(A)$ and $a' \in m^{-1}(a)$ then a' is a sequence of loops $a' = (a_0, a_1, \ldots, a_{t-1}, z_t)$ with $a_\tau \in L(z_\tau)$, or if $a = (z_0)$ then $a' = (a_0)$.[16] We define $\overline{m^{-1}}(a)$ to be the set of all such paths, so that $\overline{m^{-1}}(a) \supseteq m^{-1}(a)$. Now let $\overline{m_s^{-1}}(a)$ be the paths in $\overline{m^{-1}}(a)$ with exactly s transitions. The point is that the probability of $\overline{m_s^{-1}}(a)$ is the product of the probability of the loops and the probability of the transitions from $z_{\tau-1}$ to z_τ. This enables us to give an upper bound of order $\left[(1-C^N)^{1/N}\right]^s \epsilon^{r(A)}$.

Lemma 5. *For $a \in m(A)$ we have $P_\epsilon(\overline{m_s^{-1}}(a)|x) \leq s^{t(a)} D^{t(a)} \left[(1-C^N)^{1/N}\right]^{s-t(a)-N} \epsilon^{r(a)}$*

Proof. For any $a' = (a_0, a_1, \ldots, a_{t-1}, z_t)$ or $a' = (a_0)$ let t_τ be the number of transitions of a_τ (possibly zero). Hence an element of $\overline{m^{-1}}(a)$ starts at z_0 and either moves directly to z_1 or spends $t_1 - 1$ periods at some sequence of states that do not lie in W then reaches z_0 again, and moves directly to z_1. From $z_{\tau-1}$ it then either moves directly to z_τ or spends $t_{\tau-1} - 1$ periods at some sequence of states that do not lie in W then reaches $z_{\tau-1}$again and moves directly to z_τ. Fixing the t_τ's such a path has probability $\prod_{\tau=1}^{t(a)} P_\epsilon(z_\tau|z_{\tau-1}) P_\epsilon(t_{\tau-1}|L(z_{\tau-1}))$, and $P_\epsilon(\overline{m_s^{-1}}(a)|x)$ is the sum of these products over all $t_0, \ldots t_{t(a)} \geq 0$ which sum to $s - t(a)$:

$$P_\epsilon(\overline{m_s^{-1}}(a)|x) \quad = \textstyle\sum_{t_\tau \geq 0, t_0+t_1+\ldots+t_{t(a)}=s-t(a)} \displaystyle\prod_{\tau=1}^{t(a)} P_\epsilon(z_\tau|z_{\tau-1}) P_\epsilon(t_{\tau-1}|L(z_{\tau-1}))$$

Each product is bounded above by $D^{t(a)} \epsilon^{r(a)} \left[(1-C^N)^{1/N}\right]^{s-t(a)-N}$, and the number of terms is smaller than $(s-t(a))^{t(a)} < s^{t(a)}$. This gives the stated bound. □

To analyze A we will need loop-cutting algorithm that produces templates with resistance no smaller than $r(A)$. We say that m preserves r if $r(a) \geq r$ implies $r(m(a)) \geq r$. One such algorithm is the *r-preserving* algorithm. For any $a = (z_0, z_1, \ldots z_t)$ if no loop can be cut without reducing the resistance of a below r stop. Otherwise cut the first and shortest such loop and repeat. Observe that for an r-preserving algorithm the image $m(A)$ consists of no-zero-loop paths and is maximal since removing any loop would necessarily reduce the resistance below r. The key property of this algorithm is that it produces templates with resistance not too much bigger than r and of bounded length - and in particular that means there are finitely many templates.

Lemma 6. *For the r-preserving loop-cup algorithm $r(m(a)) \leq r + N\overline{r}$ and $t(m(a)) \leq (N-1)(r + N\overline{r})/\underline{r} \equiv t(r)$.*

Proof. Observe first that if $r(m(a)) > N\overline{r}$ then $m(a)$ must have a loop of resistance greater than zero, and hence must have a loop of resistance no greater than $N\overline{r}$. If $r(m(a)) > r + N\overline{r}$ removing such a loop leaves resistance greater than or equal to r contradicting the fact that the r-preserving algorithm can leave no such loop. To find $t(m)$ apply Lemma 1. □

[16] Since $z_t \in B \subseteq W$ then $z_\tau = z_t$ only if $\tau = 1$ or $\tau = t$. Hence if z_t is part of a loop cutting the loop results in the zero transition path (z_1).

Now let A_s be the paths in A with exactly s transitions. We now have enough tools to give an upper bound of order $\left[(1-C^N)^{1/N}\right]^s \epsilon^{r(A)}$ to the probability $P_\epsilon(A_s|x)$.

Lemma 7. $P_\epsilon(A_s|x) \le N^{t(r(A))} s^{t(r(A))} D^{t(r(A))} \left[(1-C^N)^{1/N}\right]^{s-t(r(A))-N} \epsilon^{r(A)}$.

Proof. Let m be the $r(A)$-preserving algorithm. Then for $a \in m(A)$ we have $t(a) \le t(r(A))$ and $r(a) \ge r(A)$, hence

$$\begin{aligned}
P_\epsilon(A_s|x) &\le \sum_{a\in m(A)} P_\epsilon(\overline{m_s^{-1}}(a)|x) \\
&\le \sum_{a\in m(A)} s^{t(a)} D^{t(a)} \left[(1-C^N)^{1/N}\right]^{s-t(a)-N} \epsilon^{r(a)} \\
&\le \sum_{a\in m(A)} s^{t(r(A))-1} D^{t(r(A))} \left[(1-C^N)^{1/N}\right]^{s-t(r(A))-N} \epsilon^{r(A)}
\end{aligned}$$

and there are at most $N^{t(r(A))}$ elements in $m(A)$ since the maximum length of any element is $t(r(A))$. □

Adding up over the length of the paths, we can now establish an upper bound of order $\epsilon^{r(A)}$, the same order as the lower bound in Lemma 3.

Lemma 8. *There are constants* $0 < D_1(r) < \infty$ *such that* $P_\epsilon(A|x) \le D_1(r(A))\epsilon^{r(A)}$.

Proof. We have

$$\begin{aligned}
P_\epsilon(A|x) &= \sum_{s=1}^{\infty} P_\epsilon(A_s|x) \\
&\le \sum_{s=1}^{\infty} N^{t(r(A))} s^{t(r(A))} D^{t(r(A))} \left[(1-C^N)^{1/N}\right]^{s-t(r(A))-N} \epsilon^{r(A)} \\
&= N^{t(r(A))} D^{t(r(A))} \left[(1-C^N)^{1/N}\right]^{-t(r(A))-N} \epsilon^{r(A)} \sum_{s=1}^{\infty} s^{t(r(A))} \left[(1-C^N)^{1/N}\right]^s
\end{aligned}$$

The series $\sum_{s=1}^{\infty} s^{t(r(A))} \left[(1-C^N)^{1/N}\right]^s$ is summable given that $C > 0$ (it is known as the Jonquière's function of order $-t(r(A))$). □

This completes the proof of the bounds for $P_\epsilon(A|x)$ stated in Theorem 1. For the last result, since the power to which s is raised in the sum can be arbitrary, we can use a similar argument to establish bounds for the moments $E[t^k(a)|x, A]$:

Lemma 9. *There are constants* $0 < D_2(k, r) < \infty$ *such that* $E[t^k(a)|x, A] \le D_2(k, r(A))/C^{t(A)}$.

Proof. Using Lemmas 3 and 7 we get

$$E[t^k(a)|x, A] = \frac{\sum_{s=1}^{\infty} s^k P_\epsilon(A_s|x)}{P_\epsilon(A|x)} \le \frac{\sum_{s=1}^{\infty} N^{t(r(A))} s^{t(r(A)+k} D^{t(r(A))} \left[(1-C^N)^{1/N}\right]^{s-t(r(A))-N} \epsilon^{r(A)}}{C^{t(A)}\epsilon^{r(A)}}$$

As in the previous lemma the series is summable giving the desired result. □

Appendix 2: Quasi-Direct Routes

Theorem. *[Theorem 3 in text] Let* $A = \{a|\rho(a) = \rho(\boldsymbol{A}(x, B, W))\}$ *denote the least peak resistance paths in* $\boldsymbol{A}(x, B, W) \neq \emptyset$*. Then* $\lim_{\epsilon \to 0} \frac{P_\epsilon(A|x)}{P_\epsilon(\boldsymbol{A}(x,B,W)\backslash A|x)} = \infty$.

Proof. Fix $\rho = \rho(\boldsymbol{A}(x, B, W))$. By Theorem 2 the set A consists of the set of paths of the form $a_1, a_2, \ldots, a_n, a^+$ where $a_i \in \mathcal{A}^0$ and $a^+ \in \boldsymbol{A}^+$and $r(a_i) \leq \rho, r(a^+) = \rho$. Define the set of paths $\boldsymbol{A}_{>\rho}$ as those having the form $a_1, a_2, \ldots, a_n, a_{n+1}$ where $a_i \in \mathcal{A}^0$ and $a_{n+1} \in \mathcal{A}^0 \cup \boldsymbol{A}^+$and $r(a_i) \leq \rho, r(a_{n+1}) > \rho$. We claim that $P_\epsilon(\boldsymbol{A}_{>\rho}|x) \geq P_\epsilon(\boldsymbol{A}(x, B, W)\backslash A|x)$ so that it will suffice to prove that$\lim_{\epsilon \to 0} \frac{P_\epsilon(A|x)}{P_\epsilon(\boldsymbol{A}_{>\rho}|x)} = \infty$. To see the truth of this claim, observe that if $a \in \boldsymbol{A}(x, B, W)\backslash A$ then the first part of the path must necessarily lie in $\boldsymbol{A}_{>\rho}$ so the event $\boldsymbol{A}(x, B, W)\backslash A$ implies the event $\boldsymbol{A}_{>\rho}$.

Now let $\mathcal{A}^0_\rho, \boldsymbol{A}^+_\rho, \mathcal{A}^0_{>\rho}$ denote the subsets of $\mathcal{A}^0$ and $\boldsymbol{A}^+$with resistance exactly equal to ρ and strictly bigger than ρ respectively. We compute

$$\begin{aligned}
\frac{P_\epsilon(A|x)}{P_\epsilon(\boldsymbol{A}_{>\rho}|x)} &= \frac{\sum_{n=0}^\infty P^n_\epsilon(\mathcal{A}^0_\rho|x) P_\epsilon(\boldsymbol{A}^+_\rho|x)}{\sum_{n=0}^\infty P^n_\epsilon(\mathcal{A}^0_\rho|x) P_\epsilon(\mathcal{A}^0_{>\rho} \cup \boldsymbol{A}^+_{>\rho}|x)} \\
&= \frac{P_\epsilon(\boldsymbol{A}^+_\rho|x)}{P_\epsilon(\mathcal{A}^0_{>\rho} \cup \boldsymbol{A}^+_{>\rho}|x)} \geq \frac{P_\epsilon(\boldsymbol{A}^+_\rho|x)}{P_\epsilon(\mathcal{A}^0_{>\rho}|x) + P_\epsilon(\boldsymbol{A}^+_{>\rho}|x)}
\end{aligned}$$

and the result now follows directly from Theorem 1 on the probability of direct paths. □

When $B = W$ the decomposition also makes it easy to do computations since the loops a_i are independent identically distributed random variables. Specifically, for $f : \mathcal{A}^0 \to \Re$ and $a^- \in A^-$ define $f^-(a^-) \equiv \sum_{i=1}^{n(a)} f(a_i)$. Then for any function $g(n)$ of the number of loops we have

Lemma 10. *if* $B = W$ *then* $E(f^- g|x, \boldsymbol{A}^-) = E(f|x, \mathcal{A}^0)E(ng|x, \boldsymbol{A}^-)$*, and* $E(n|x, \boldsymbol{A}^-) = 1/P_\epsilon(\boldsymbol{A}^+|x)$

Proof. Since $B = W$ we have $P_\epsilon(\mathcal{A}^0|x) + P_\epsilon(\boldsymbol{A}^+|x) = 1$, while $\mathcal{A}^0$ and $\boldsymbol{A}^+$ are disjoint. Then

$$E(f^- g|x, A^-) = E[\sum_{i=1}^n f(a_i)g|x, A^-] = E[\sum_{i=1}^n E[f(a_i)g|x, A^-, n]|x, A^-] = E[\sum_{i=1}^n gE[f(a_i)|x, A^-, n]|x, A^-].$$

The event $(\boldsymbol{A}^-, n)$ is exactly the event $a_i \in A^0$ for $i = 1, 2, \ldots, n$ and $a_{n+1} \in \boldsymbol{A}^+$ and conditional on x these are independent events. Hence $E(f(a_i)|x, \boldsymbol{A}^-, n) = E(f|x, \mathcal{A}^0)$.We conclude that

$$\begin{aligned}
E(f^- g|x, \boldsymbol{A}^-) &= E[\sum_{i=1}^n gE(f|x, \mathcal{A}^0)|x, \boldsymbol{A}^-] \\
&= E(f|x, \mathcal{A}^0)E[\sum_{i=1}^n g|x, \boldsymbol{A}^-] = E(f|x, \mathcal{A}^0)E[ng|x, \boldsymbol{A}^-].
\end{aligned}$$

This is the first result. Also since $P_\epsilon(\mathcal{A}^0|x) + P_\epsilon(\boldsymbol{A}^+|x) = 1$ it follows that n is geometrically distributed with success probability $P_\epsilon(\boldsymbol{A}^+|x)$ which gives the stated expected value. □

Recall that $a^- \in \boldsymbol{A}^-$ is a sequence $a_1, a_2, \ldots, a_n$ with $a_i \in \mathcal{A}^0$ loops at x. Now let $M(A^0)$ be the number of loops that lie in $A^0 \subseteq \mathcal{A}^0$. That is, if $f : \mathcal{A}^0 \to \Re$ is the indicator of A^0 ($f(a^0) = 1$ if $a^0 \in A^0$ and $f(a^0) = 0$ for $a^0 \in \mathcal{A}^0 \setminus A^0$) then $M(A^0)$ is the aggregate f^-. Also, let t_- be the amount of time along a^- spent outside of $\Omega(x)$.

Theorem. *[Theorem 4 in text] If $B = W$ we have for some $C_1, D_3 > 0$*

$$1 + C_1 \epsilon^{-r(\boldsymbol{A}^+)} \leq E[t|x, \boldsymbol{A}(x, B, W)] \leq D_2(1,0)C^{-N} + D_3 C^{-2N} \epsilon^{-r(\boldsymbol{A}^+)}$$

and for $A^0 \subseteq \mathcal{A}^0$

$$C_1 C^{t(A^0)} \epsilon^{r(A^0) - r(\boldsymbol{A}^+)} \leq E[M(A^0)|x, \boldsymbol{A}^-] \leq D_1(r(A^0)) C^{-2N} \epsilon^{r(A^0) - r(\boldsymbol{A}^+)}.$$

Moreover, $\lim_{\epsilon \to 0} E[t_-/t|x, \boldsymbol{A}^-] = 0$.

Proof. From Lemma 10 $E[t|x, \boldsymbol{A}^-] = E[t|x, \mathcal{A}^0]/P_\epsilon(\boldsymbol{A}^+|x)$. Moreover, recalling that $t(A^0)$ is the number of transitions in the shortest of the least resistance paths in A^0 and that since $\mathcal{A}^0$ contains all zero resistance loops $r(\mathcal{A}^0) = 0$ and the shortest of these loops is no longer than N so $t(\mathcal{A}^0) \leq N$; analogously, $\boldsymbol{A}^+$ contains templates without loops hence $t(\boldsymbol{A}^+) \leq N$ and consequently $r(\boldsymbol{A}^+) \leq N\overline{r}$. Hence from Theorem 1 we have $1 \leq E[t|x, \boldsymbol{A}^+] \leq D_2(1, r(\boldsymbol{A}^+))/C^N \leq \max_{r \leq N\overline{r}} D_2(1,r)/C^N$, $1 \leq E[t|x, \mathcal{A}^0] \leq D_2(1,0)/C^N$ and $C^N \epsilon^{r(\boldsymbol{A}^+)} \leq P_\epsilon(\boldsymbol{A}^+|x) \leq D_1(r(\boldsymbol{A}^+))\epsilon^{r(\boldsymbol{A}^+)} \leq \max_{r \leq N\overline{r}} D_1(r)\epsilon^{r(\boldsymbol{A}^+)}$. This gives the stated bound on $E[t|x, \boldsymbol{A}(x, B, W)]$.

Next, given $A^0 \subseteq \mathcal{A}^0$ and f the indicator of A^0, Lemma 10 gives $E[M(A^0)|x, \boldsymbol{A}^-] = E[f^-|x, \boldsymbol{A}^-] = E(f|x, \mathcal{A}^0)/P_\epsilon(\boldsymbol{A}^+|x) = P_\epsilon(A^0|x, \mathcal{A}^0)/P_\epsilon(\boldsymbol{A}^+|x) = P_\epsilon(A^0|x)/[P_\epsilon(\mathcal{A}^0|x)P_\epsilon(\boldsymbol{A}^+|x)]$. Applying Theorem 1 then gives

$$[D_1(0)D_1(r(\boldsymbol{A}^+))]^{-1} C^{t(A^0)} \epsilon^{r(A^0) - r(\boldsymbol{A}^+)} \leq E[M(A^0)|x, \boldsymbol{A}^-] \leq D_1(r(A^0)) C^{-2N} \epsilon^{r(A^0) - r(\boldsymbol{A}^+)}.$$

Again making use of $r(\boldsymbol{A}^+) \leq N\overline{r}$ we have the stated bound on $E[M(A^0)|x, \boldsymbol{A}^-]$.

Finally by Lemma 10 $E[t_-/t|x, \boldsymbol{A}^-] \leq E[t_-/n|x, \boldsymbol{A}^-] = E[t_-|x, \mathcal{A}^0]$. Now split $\mathcal{A}^0$ into two disjoint sets $\mathcal{A}^0_0$ of paths of zero resistance and $\mathcal{A}^0_r$ of positive resistance, where r is the least positive resistance in $\mathcal{A}^0$. Then $E[t_-|x, \mathcal{A}^0] = E[t_-|x, \mathcal{A}^0_0]P_\epsilon[\mathcal{A}^0_0|x, \mathcal{A}^0] + E[t_-|x, \mathcal{A}^0_r]P_\epsilon[\mathcal{A}^0_r|x, \mathcal{A}^0]$. However $E[t_-|x, \mathcal{A}^0_0] = 0$ by definition, while by Theorem 1

$$E[t_-|x, \mathcal{A}^0_r]P_\epsilon[\mathcal{A}^0_r|x, \mathcal{A}^0] \leq [D_2(1,r)/C^{t(\mathcal{A}^0_r)}]D_1(r)\epsilon^r \to 0$$

. ∎

Appendix 3: Ergodic Probabilities and Bounds

Theorem. *[Theorem 6 in text] If $y \in \Omega(x)$ then*

$$\lim_{\epsilon \to 0} \frac{\mu_\epsilon(x)}{\mu_\epsilon(y)} = \frac{\overline{\mu}_0(x)}{\overline{\mu}_0(y)}.$$

Proof. Partition the matrix P_ϵ with rows corresponding to source states and columns to target states into P^{ij}_ϵ where $i,j = 1$ corresponds to $\Omega(x)$ and $i,j = 2$ corresponds to $\Omega \backslash Z$. In particular P^{11} is square, the size of $\Omega(x)$. Correspondingly let e^i be the column vectors of ones with length corresponding to $i = 1,2$. Define the row vector $\overline{\mu}_\epsilon(z) = \mu_\epsilon(z)/\sum_{y \in \Omega(x)} \mu_\epsilon(y)$, and partition this vector conformally. Since $\overline{\mu}_\epsilon$is normalized to one on $\Omega(x)$ and $\overline{\mu}_0$ is strictly positive, it suffices to prove that as $\epsilon \to 0$ every limit point $\overline{\mu}^1_\epsilon$ is equal to $\overline{\mu}^1_0$ where we include the superscript to emphasize that we are dealing only with the invariant distribution on $\Omega(x)$. The invariance condition is $\overline{\mu}^1_\epsilon = \overline{\mu}^1_\epsilon P^{11}_\epsilon + \overline{\mu}^2_\epsilon P^{21}_\epsilon$. Multiplying this on the right by e^1 we get $1 = \overline{\mu}^1_\epsilon P^{11}_\epsilon e^1 + \overline{\mu}^2_\epsilon P^{21}_\epsilon e^1$ while the fact that P_ϵ is a Markov kernel means that $P^{11}_\epsilon e^1 + P^{12}_\epsilon e^2 = e^1$ or $P^{11}_\epsilon e^1 = e^1 - P^{12}_\epsilon e^2$. Substituting we see that $1 = \overline{\mu}^1_\epsilon(e^1 - P^{12}_\epsilon e^2) + \overline{\mu}^2_\epsilon P^{21}_\epsilon e^1 = 1 - \overline{\mu}^1_\epsilon P^{12}_\epsilon e^2 + \overline{\mu}^2_\epsilon P^{21}_\epsilon e^1$ so that $\overline{\mu}^2_\epsilon P^{21}_\epsilon e^1 = \overline{\mu}^1_\epsilon P^{12}_\epsilon e^2$, which says roughly that the steady state flow into $\Omega(x)$ must equal the steady state flow out. As $\epsilon \to 0$ $P_\epsilon^{12} \to 0$ since these are the probabilities of leaving the irreducible set $\Omega(x)$, it follows that $\overline{\mu}_\epsilon^2 P_\epsilon^{21} e^1 \to 0$. But $\overline{\mu}_\epsilon^2 P_\epsilon^{21}$ is a non-negative vector, so $\overline{\mu}_\epsilon^2 P_\epsilon^{21} e^1 \to 0$ is possible only if $\overline{\mu}_\epsilon^2 P_\epsilon^{21} \to 0$. Then in the invariance condition $\overline{\mu}_\epsilon^1 = \overline{\mu}_\epsilon^1 P_\epsilon^{11} + \overline{\mu}_\epsilon^2 P_\epsilon^{21}$ as $P_\epsilon^{11} \to P_0^{11}$ if $\overline{\mu}_{00}^1$ is a limit point of $\overline{\mu}_\epsilon^1$ it must satisfy the limiting condition that $\overline{\mu}_{00}^1 = \overline{\mu}_{00}^1 P_0^{11}$. However, as $\Omega(x)$ is irreducible this equation has only one solution $\overline{\mu}_0^1$, so we conclude that in fact $\overline{\mu}_\epsilon^1 \to \overline{\mu}_0^1$. □

Recall that $\mathscr{W}$ are all the irreducible classes and that $r(\Omega(x)) = \min_{\Omega(y)\in\mathscr{W}} r(\Omega(x), \Omega(y))$.

Theorem. *[Theorem 7 in text] Allowing that $\Omega(x)$ may be empty, if $A = \mathcal{A}(x, \{y\}, \{x\} \cup \{y\} \cup (\Omega \backslash \Omega(x))$ are the direct routes from x to y with forbidden set $\{x\} \cup \{y\} \cup (\Omega\backslash\Omega(x))$ then $\mu_\epsilon(y) \geq \mu_\epsilon(x) C^N \epsilon^{r(A)}$. There is also a constant D_4 such that if $x \in \Omega(x)$ and there is a zero resistance path from y to x then also $\mu_\epsilon(y) \leq \mu_\epsilon(x) D_4 \epsilon^{\min\{r(A), r(\Omega(x))\}}$.*

Proof. We use the standard fact about Markov ergodic probabilities as used for example by Ellison (2000): if we let $N_\epsilon(y, x|x)$ be the expected number of times y occurs before x starting at x then $\mu_\epsilon(y) = \mu_\epsilon(x) N_\epsilon(y, x, |x)$.

The lower bound is immediate: since with probability $P_\epsilon(A)$ we have y hit once without returning to x we have from Theorem 1 $\mu_\epsilon(y) = \mu_\epsilon(x) N_\epsilon(y, x|x) \geq \mu_\epsilon(x) P_\epsilon(A) \geq \mu_\epsilon(x) C^N \epsilon^{r(A)}$.

Next we suppose that y has zero resistance for getting to $x \in \Omega(x)$. We use the reverse condition $\mu_\epsilon(x) = \mu_\epsilon(y) N_\epsilon(x, y, |y)$, so we must find a lower bound on $N_\epsilon(x, y|y)$. Let $A_1 = \mathcal{A}(y, \{x\}, \{x\} \cup \{y\} \cup (\Omega\backslash\Omega(x)))$. Observe that $N_\epsilon(x, y|y) \geq P_\epsilon(A_1) N_\epsilon(x, y|x)$. Since there is a zero resistance path from y to x we have from Theorem 1 the bound $P_\epsilon(A_1) \geq C^N$, so $N_\epsilon(x, y|y) \geq C^N N_\epsilon(x, y|x)$.

Now define set $B = \{y\} \cup (\Omega\backslash\Omega(x))$ and $A_2 = \mathcal{A}(x, B, \{x\} \cup B)$. Then $N_\epsilon(x, y|x) \geq N_\epsilon(x, B|x))$. Since starting at x B and $\sim B = A_2$ are mutually exclusive independent events, $N_\epsilon(x, B|x)) = 1/P_\epsilon(\sim B) = 1/P_\epsilon(A_2)$. From Theorem 1 $P_\epsilon(A_2) \leq D_5 \epsilon^{r(A_2)}$, and we get $N_\epsilon(x, y|y) \geq C^N \epsilon^{-r(A_2)}/D_5$, or $\mu_\epsilon(y) \leq \mu_\epsilon(x) D_4 \epsilon^{r(A_2)}$.

Finally the event A_2 is contained in the event $\mathcal{A}(x, \{y\}, \{x\} \cup \{y\} \cup (\Omega\backslash\Omega(x))) \cup \mathcal{A}(x, (\Omega\backslash\Omega(x))), \{x\} \cup (\Omega\backslash\Omega(x)))$. Hence

$$r(A_2) = \min\{r\left(\mathcal{A}(x, \{y\}, \{x\} \cup \{y\} \cup (\Omega\backslash\Omega(x)))\right), r\left(\mathcal{A}(x, (\Omega\backslash\Omega(x))), \{x\} \cup (\Omega\backslash\Omega(x)))\right)\}.$$

However $r\left(\mathcal{A}(x, \Omega\backslash\Omega(x), \{x\} \cup (\Omega\backslash\Omega(x)))\right\} = r(\Omega(x))$ and $r\left(\mathcal{A}(x, \{y\}, \{x\} \cup \{y\} \cup (\Omega\backslash\Omega(x)))\right) = r(A)$ which gives the desired upper bound. □

Appendix 4: Ergodic Probabilities and Circuits

We are given a finite set of nodes $\psi \in \Psi$ and a resistance function $r(\psi, \phi)$. For any $\psi \in \Psi$ we define the *least resistance* $r(\psi) = \min_{\phi\in\Psi} r(\psi, \phi)$. We are interested in trees T on Ψ. For any such tree and any ψ let $T(\psi)$ denote the unique predecessor of ψ on the tree (which is null for the unique root). Note that we follow the standard terminology that the predecessor is closer to the root - in contrast to Young who follows the logic of the Markov process in imagining that the node closer to the root is the successor node. The *resistance of the tree* T is defined to be $r(T) = \sum_{\psi\in\Psi} r(\psi, T(\psi))$ where $r(\psi, \emptyset) \equiv 0$.

Our goal is to characterize least resistance trees by showing how they are constructed out of groups of nodes that we call circuits. Say that a subset $\Phi \subseteq \Psi$ is a *circuit* if for each pair $\phi_1, \psi \in \Phi$ there is a path $\phi_1, \phi_2, \ldots, \phi_n \in \Phi$ with $\phi_n = \psi$ such that for $\tau = 2, 3, \ldots n$ we have

$r(\phi_{\tau-1}, \phi_\tau) = r(\phi_{\tau-1})$, that is, there is a path from $\phi_1$to ψ in Φ such that each connection has least resistance. We say that a circuit Φ is *consolidated within the tree* T if there is a $\psi \in \Phi$ and for all $\phi \in \Phi$, $\phi \neq \psi$ we have $T(\phi) \in \Phi$ and $r(\phi, T(\phi)) = r(\phi)$. In other words, in the consolidated tree the circuit Φ forms a subtree with root ψ, and each connection within the circuit has least resistance. We refer to ψ as the *top of the circuit.*

Intuitively if we think of the circuit as a circle of least resistance connections then we will break that circle after ψ to make a subtree and use ψ to connect this subtree to the the rest of the tree. Breaking the connection saves at least $r(\psi)$, while making the new connection costs $r(\psi, T(\psi))$, hence we define the *modified resistance from* ψ *to* ϕ as $R(\psi, \phi) = r(\psi, \phi) - r(\psi)$.

In the next lemma we consolidate a circuit within a tree by breaking it after the node that minimizes modified resistance. By so doing, the resistance of the tree cannot increase. The intuition is that the system leaves the circuit most likely through such a node. Indeed, the system can remain within Φ with least resistance transitions, so given a fixed target $\phi \notin \Phi$, from any node $\psi \in \Phi$ the system can remain in Φ with resistance $r(\psi)$ or go to ϕ with resistance $r(\psi, \phi) \geq r(\psi)$. Since the probability of leaving ψ is of order $\epsilon^{r(\psi)}$ the expected length of any visit to ψ is $1/\epsilon^{r(\psi)}$; on the other hand the probability of going to ϕ is of order $\epsilon^{r(\psi,\phi)}$, so the probability of going to ϕ during a visit in ψ is of order $(1/\epsilon^{r(\psi)})\epsilon^{r(\psi,\phi)}$; the expected number of visits to ψ for this to occur with high probability (probability one) is then k such that $k(1/\epsilon^{r(\psi)})\epsilon^{r(\psi,\phi)} = 1$, that is $1/\epsilon^{r(\psi,\phi)-r(\psi)} = 1/\epsilon^{R(\psi,\phi)}$. This number of visits before exit is smallest for the ψ which has minimum $R(\psi, \phi)$ over $\psi \in \Phi$.

Lemma 11. *Suppose that T has root ψ and that Φ is a circuit. Then there is a tree T' with root ψ such that $r(T') \leq r(T)$ and Φ is consolidated in T' with the additional properties that (1) if $\phi' \notin \Phi$ then $T'(\phi') = T(\phi')$ and (2) if ϕ is the top of Φ in T' then $R(\phi, T'(\phi)) = \min_{\phi' \in \Phi} R(\phi', T'(\phi))$.*

Proof. Let T have root ψ and let $\phi^* \in \Phi$ be such that the unique path from ϕ^* to the root ψ contains no element of Φ. If $\phi^* = \psi$ take $\phi = \phi^*$. Otherwise choose as top a $\phi \in \Phi$ such that $r(\phi, T(\phi^*)) - r(\phi) = \min_{\phi' \in \Phi} r(\phi', T(\phi^*)) - r(\phi')$. We now use tree surgery to create a sequence of new trees ending in the desired tree T'. As we proceed we never cut a connection originating in any set other than Φ so that property (1) will be satisfied.

At each step Φ will be divided into two sets $\Phi_\phi, \Phi_{\sim\phi} = \Phi \backslash \Phi_\phi$. The first set Φ_ϕ will contain at least ϕ and consist of those elements of Φ that are already consolidated with ϕ at the top, and such that no element of $\Phi_{\sim\phi}$ appears between ϕ and the root. We will proceed constructing new trees by moving one element from $\Phi_{\sim\phi}$ to Φ_ϕ making sure that all properties are preserved.

We start the process. If $\phi = \psi$ or $\phi = \phi^*$ we do nothing. Otherwise cut ϕ from the tree and paste it to $T(\phi^*)$. Observe that this increased the modified resistance of the tree by at most $r(\phi, T(\phi^*)) - r(\phi)$. Let Φ_ϕ be the maximal set consolidated with ϕ at the top: this set now contains at least ϕ.

We now continue the process until $\Phi_{\sim\phi}$ is empty. Pick an element $\phi' \in \Phi_{\sim\phi}$. Because Φ is a circuit there is a least modified resistance path in Φ from ϕ' to ϕ. Let ϕ_τ be the last element in $\Phi_{\sim\phi}$ that is reached on this path. Then cut ϕ_τ from the tree and paste it to $\phi_{\tau+1}$. Notice that this cannot increase the modified resistance of the tree since the connection from ϕ_τ to $\phi_{\tau+1}$ has least modified resistance. Moreover, if $\phi \neq \phi^*$ then at some step $\phi_\tau = \phi^*$ and at this step the resistance of the tree is decreased by exactly $r(\phi^*, T(\phi^*)) - r(\phi^*)$. Once again let Φ_ϕ be the maximal set consolidated with ϕ at the top: this set now contains at least one more element ϕ_τ.

When we are finished we end up with the new tree T'. Now observe that either $\phi = \phi^*$ or the resistance over the original tree was increased only in the first step, by at most $r(\phi, T(\phi^*)) - r(\phi)$, and it was decreased by $r(\phi*, T(\phi^*)) - r(\phi^*)$ when we pasted ϕ^*. Since by the choice of ϕ we have $r(\phi, T(\phi^*)) - r(\phi) \leq r(\phi^*, T(\phi^*)) - r(\phi^*)$. In all other cases the resistance did not increase so that $r(T') \leq r(T)$. Since by construction $T'(\phi) = T(\phi^*)$ we have $R(\phi, T'(\phi)) = \min_{\phi' \in \Phi} R(\phi', T'(\phi))$. □

We now focus on least resistance trees. Let $\mathscr{T}(\psi)$ be the set of trees with root ψ, $r_\psi = \min_{T \in \mathscr{T}(\psi)} r(T)$ be the least resistance of any tree with root ψ and $\mathscr{T}_\psi = \arg\min_{T \in \mathscr{T}(\psi)} r(T)$ be the set of least resistance trees with root ψ.

Lemma 12. *If ψ, ϕ are in the same circuit then $r_\psi - r_\phi = r(\phi) - r(\psi)$.*

Proof. Suppose $\psi, \phi \in \Phi$ where Φ is a circuit. Then we can choose a path $\phi_1, \ldots, \phi_\nu, \ldots, \phi_n \in \Phi$ with $\phi_1 = \psi, \phi_\nu = \phi, \phi_n = \psi$ such that for $\tau = 2, 3, \ldots n$ we have $r(\phi_{\tau-1}, \phi_\tau) = r(\phi_{\tau-1})$. Choose $T_1 \in \mathscr{T}_{\phi_1}$, and supposing that $T_{\tau-1}$ has root $\phi_{\tau-1}$ define T_τ as the tree in which we cut ϕ_τ from $T_{\tau-1}$, make it the root of T_τ and paste the root of $T_{\tau-1}$ to ϕ_τ. This tree has root ϕ_τ and resistance $r(T_\tau) \leq r(T_{\tau-1}) + r(\phi_{\tau-1}, \phi_\tau) - r(\phi_\tau) = r(T_{\tau-1}) + r(\phi_{\tau-1}) - r(\phi_\tau)$. Hence $r(T_\tau) \leq r(T_1) + r(\phi_1) - r(\phi_\tau)$. Since $\phi_n = \phi_1$, we conclude that $r(T_n) \leq r(T_1)$ and since T_1 had least resistance, it must be that $r(T_n) = r(T_1)$. Hence all the inequalities must hold with equality, that is, $r(T_\tau) = r(T_1) + r(\phi_1) - r(\phi_\tau)$. Choosing $\tau = \nu$ we then have $r(T_\tau) = r_\psi + r(\psi) - r(\phi)$, whence $r_\phi \leq r_\psi + r(\psi) - r(\phi)$; but by interchanging ϕ and ψ and rearranging we get $r_\phi \geq r_\psi + r(\psi) - r(\phi)$, therefore $r(T_\tau) = r_\phi$; this gives the conclusion. □

We now assume that for $\epsilon > 0$ P_ϵ is ergodic so that there is a unique ergodic probability distribution μ_ϵ on the state space Z. Let $\mathscr{T}_S(\psi)$ denote all trees over a set S with root ψ and set

$$M_\epsilon(x) = \sum_{T \in \mathscr{T}_Z(x)} \prod_{x \in Z} P_\epsilon(T(x)|x).$$

Following Young and Freidlin and Wentzell we observe

$$\mu_\epsilon(x) = \frac{M_\epsilon(x)}{\sum_{z \in Z} M_\epsilon(z)}.$$

Let the resistance $r(x, y)$ on Z be the ordinary resistance. Observing from Cayley's formula that N^{N-2} is the number of trees with the same root over N nodes it follows that

Lemma 13. *The ratio of ergodic probabilities satisfies the bounds*

$$\frac{C^N}{N^{N-2} D^N} \epsilon^{r_x - r_y} \leq \frac{\mu_\epsilon(x)}{\mu_\epsilon(y)} \leq \frac{N^{N-2} D^N}{C^N} \epsilon^{r_x - r_y}.$$

Proof. Recall the bounds $C\epsilon^{r(x,z)} \leq P_\epsilon(z|x) \leq D\epsilon^{r(x,z)}$ on transition probabilities. Observe that a sum of reals is bounded below by the largest of them and above by the largest times the number of terms. The given bounds on the ratio $\mu_\epsilon(x)/\mu_\epsilon(y) = M_\epsilon(x)/M_\epsilon(y)$ then follows. □

These bounds are in terms of resistances of least resistance trees. The next goal is to translate them in terms of appropriate resistances of least resistance paths. Take as Ψ the set $\mathscr{W}$ of irreducible classes in P_0, and define the resistance $r(\psi, \phi)$ for $\psi, \phi \in \mathscr{W}$ as the least resistance of any direct path from $x \in \psi$ to the target ϕ with forbidden set $W = \{x\} \cup \Omega \setminus \Omega(x)$ - paths which can stay in $\Omega(x)$ as long as they do not go back to x, and when they leave it they do not touch other irreducible sets. Note that $r(\psi, \phi)$ is independent of the particular choice of x in ψ since there are zero resistance paths from any point in ψ to any other. As before $r(\psi) = \min_{\phi \in \Psi} r(\psi, \phi)$.

Applying Lemma 12 give as immediate corollary

Theorem. *[Theorem 8 in text] If the irreducible classes $\Omega(x)$ and $\Omega(y)$ are in the same circuit then*

$$\frac{C^N}{N^{N-2}D^N}\epsilon^{r(\Omega(y))-r(\Omega(x))} \leq \frac{\mu_\epsilon(x)}{\mu_\epsilon(y)} \leq \frac{N^{N-2}D^N}{C^N}\epsilon^{r(\Omega(y))-r(\Omega(x))}.$$

This goes one step in the desired direction but applies only to elements of a given circuit. In general, we can find least the least resistance of trees in Z by finding the least resistance of trees in $\mathscr{W}$.

Lemma 14. *If $x \in \psi \in \mathscr{W}$ then $r_x = r_\psi$.*

Proof. Young proves this theorem (Lemma 2 in his Appendix) for the case where the resistance, call it $r^*(\psi, \phi)$, is the least resistance of any path from ψ to ϕ - that is, he allows the path to pass through irreducible sets ϕ' which are neither ψ nor ϕ. (Ellison does the same in his definition of the modified co-radius.) Our resistance is in general larger than Young's since we do not allow paths to pass through these other irreducible sets. However, his proof requires only minor modification to yield the stronger result. Young first shows that the least resistance r^*_ψ of any tree on $\mathscr{W}$ with root ψ is greater than or equal to r_x. Since $r_\psi \geq r^*_\phi$ we have the immediate implication that $r_\psi \geq r_x$.

The second part of Young's proof shows that $r^*_\psi \leq r_x$. He proceeds by showing how to reorganize a least resistance tree $T \in \mathscr{T}_x$ on Z into a tree over $\mathscr{W}$ with root ψ that has no greater resistance than T. The easiest way to do this would be by simply taking one point from each irreducible class and using the resistance between those points to get a tree over $\mathscr{W}$. However, this does not work because there can be double-counting if paths in T join between irreducible classes. Young shows how to avoid double-counting by reorganizing the tree. We can use his construction if we can avoid having or creating paths between irreducible classes that contain elements of a third irreducible class. This is the case if we start by choosing the "right" least resistance tree and the "right" point from each irreducible class before we apply Young's method.

Observe that each irreducible class is a circuit, so by Lemma 11 we can consolidate all the irreducible classes into a tree T' which has no greater resistance, hence is also least resistance. So start Young's construction with the tree T'. The first step in Young's proof is to choose one point from each irreducible class - these are what Young calls special vertices. We do this by choosing from each irreducible class the top of the corresponding circuit in the tree T'. Observe that because the tree is consolidated the path from any special vertex to the next special vertex y in the direction of the root cannot contain elements of any irreducible class other than $\Omega(y)$.

Now apply Young's construction to eliminate junctions. Observe that when Young cuts subtrees from a vertex y that is not in an irreducible set this preserves the consolidated structure of irreducible sets: those irreducible sets that lie further from the root than y in the tree are necessarily entirely contained in the subtree. Consequently we never need to cut at junctions y that are in irreducible sets: since the tree is consolidated the path in the tree from y to the top of the circuit (the special vertex) has zero resistance and no double-counting is involved.

Finally, when Young pastes cuts from the junction y back into the tree he implicitly introduces new path segments starting at y and ending at a special vertex z that has zero resistance from y.

However, these implicit paths cannot contain elements of any irreducible set ϕ other than $\Omega(z)$. If they did the path could not have zero resistance since there is no path from $\phi \neq \Omega(z)$ to $\Omega(z)$ that has zero resistance. Hence at the end of Young's procedure we find that the paths along which resistance is computed - those from one special vertex to the next special vertex in the direction of the root - do not contain a vertex from a third irreducible class. Therefore $r_\psi \leq r_x$. $\square$

Our next goal is to recursively compute r_ϕ and by doing so find bounds on $\mu_\epsilon(x)/\mu_\epsilon(y)$ - without the restriction that $\Omega(x)$ and $\Omega(y)$ be in the same circuit. We suppose we are given a set Ψ^{k-1} with resistance function $r^{k-1}(\psi^{k-1}, \phi^{k-1})$ and a partition of Ψ^{k-1} into circuits $\psi^k \in \Psi^k$. As we observed already this means considering a "higher order" time horizon. We will see that what happens in all relevant time spans matters. Note that we will take $\Psi^0 = \mathscr{W}$ (and not Z), so an element $\psi^1 \in \Psi^1$ will be a circuit of irreducible sets. We define the modified resistance function $R^{k-1}(\psi^{k-1}, \phi^{k-1}) = r^{k-1}(\psi^{k-1}, \phi^{k-1}) - r^{k-1}(\psi^{k-1})$, and we define a resistance function on Ψ^k by the least modified resistance: $r^k(\psi^k, \phi^k) = \min_{\psi^{k-1} \in \psi^k, \phi^{k-1} \in \phi^k} R^{k-1}(\psi^{k-1}, \phi^{k-1})$. Recall from the discussion on page 26 that this is actually the most likely route from ψ^k to ϕ^k. Then the following formula holds, where notice that the term $\sum_{\phi^{k-1} \in \Psi^{k-1}} r^{k-1}(\phi^{k-1})$ is a constant independent of the tree in question.

Lemma 15. *If* $\psi^{k-1} \in \psi^k$ *then* $r^{k-1}_{\psi^{k-1}} = r^k_{\psi^k} - r^{k-1}(\psi^{k-1}) + \sum_{\phi^{k-1} \in \Psi^{k-1}} r^{k-1}(\phi^{k-1})$.

Proof. Suppose we have a tree T^{k-1} on Ψ^{k-1} that is consolidated with respect to all the circuits in Ψ^k, and let ψ^{k-1} be its root. The fact that T^{k-1} is consolidated means that the top of each circuit has a predecessor which belongs to a different circuit. Let $t(T^{k-1}, \psi^k) \in \Psi^{k-1}$ denote the top of circuit ψ^k in T^{k-1}. Then if $T^{k-1}(t(T^{k-1}, \psi^k)) = \phi^{k-1} \in \phi^k \neq \psi^k$ (where if ϕ^{k-1} is null we set $\phi^k = \emptyset$ as well), we may define $T^k(\psi^k) = \phi^k$. In this way we define a tree on Ψ^k. We have $r^{k-1}(T^{k-1}) = \sum_{\phi^{k-1} \in \Psi^{k-1}} r^{k-1}(\phi^{k-1}, T^{k-1}(\phi^{k-1}))$. However, since the tree is consolidated, for any ϕ^{k-1} not at the top of the corresponding circuit ϕ^k we have $r^{k-1}(\phi^{k-1}, T^{k-1}(\phi^{k-1})) = r^{k-1}(\phi^{k-1})$, hence we may write

$$r^{k-1}(T^{k-1}) = \sum_{\phi^{k-1} \in \Psi^{k-1}} r^{k-1}(\phi^{k-1}) - r^{k-1}(\psi^{k-1}) + \sum_{\phi^k \in \Psi^k} \rho^{k-1}(t(T^{k-1}, \phi^k), T^{k-1}(t(T^{k-1}, \phi^k))).$$

Now start with a least resistance tree $T^{k-1} \in \mathscr{T}_{\psi^{k-1}}$. By Lemma 11 we may consolidate this with respect to all the circuits in Φ^k to get another least resistance tree $\tilde{T}^{k-1} \in \mathscr{T}_{\psi^{k-1}}$. By the previous computation and the definition of r^k we see that

$$\begin{aligned} r^{k-1}_{\psi^{k-1}} = r^{k-1}(\tilde{T}^{k-1}) &= \sum_{\phi^{k-1} \in \Psi^{k-1}} r^{k-1}(\phi^{k-1}) - r^{k-1}(\psi^{k-1}) + \sum_{\phi^k \in \Psi^k} \rho^{k-1}(t(T^{k-1}, \phi^k), T^{k-1}(t(T^{k-1}, \phi^k))) \\ &\geq \sum_{\phi^{k-1} \in \Psi^{k-1}} r^{k-1}(\phi^{k-1}) - r^{k-1}(\psi^{k-1}) + \sum_{\phi^k \in \Psi^k} r^k(\phi^k, T^k(\phi^k)) \\ &\geq \sum_{\phi^{k-1} \in \Psi^{k-1}} r^{k-1}(\phi^{k-1}) - r^{k-1}(\psi^{k-1}) + r^k_{\psi^k}. \end{aligned}$$

Next start with a least resistance tree $T^k \in \mathscr{T}_{\psi^k}$, where $\psi^{k-1} \in \psi^k$, and construct a tree on Ψ^{k-1} as follows. For the root $\phi^k = \psi^k$ define $\phi^{k-1} = \psi^{k-1}$. For given non-root ϕ^k and $T^k(\phi^k)$ there are points $\phi^{k-1} \in \phi^k$ and $\tilde{\phi}^{k-1} \in T^k(\phi^k)$ such that $r^k(\phi^k, T^k(\phi^k)) = r(\phi^{k-1}, \tilde{\phi}^{k-1}) - r(\phi^{k-1})$. For each ϕ^k consolidate the tree over ϕ^k with root ϕ^{k-1} to get a tree $T[\phi^k, \phi^{k-1}]$. Now define a tree on Ψ^{k-1} by putting together these subtrees as follows: if $\hat{\phi}^{k-1}$ is in $T[\phi^k, \phi^{k-1}]$ but is not the root, set $T^{k-1}(\hat{\phi}^{k-1}) = T[\phi^k, \phi^{k-1}](\hat{\phi}^{k-1})$. For the root ϕ^{k-1} set $T^{k-1}(\hat{\phi}^{k-1}) = \tilde{\phi}^{k-1}$. This is clearly a tree with root ψ^{k-1}, and we see that the resistance is

$$\begin{aligned} r^{k-1}_{\psi^{k-1}} \leq r^{k-1}(T^{k-1}) &= \sum_{\phi^{k-1} \in \Psi^{k-1}} r^{k-1}(\phi^{k-1}) - r^{k-1}(\psi^{k-1}) + \sum_{\phi^k \in \Psi^k} r^k(\phi^k, T^k(\phi^k)) \\ &= \sum_{\phi^{k-1} \in \Psi^{k-1}} r^{k-1}(\phi^{k-1}) - r^{k-1}(\psi^{k-1}) + r^k_{\psi^k}. \end{aligned}$$

Putting together the two inequalities gives the desired result. $\square$

Lemma 16. *If Ψ^k has at least two elements is has at least one non-trivial circuit.*

Proof. Starting at an arbitrary point $\psi^k \in \Psi^k$ choose a path of least resistance. Since Ψ^k is finite, this must eventually have a loop, and that loop is necessarily a circuit. $\square$

We can now recursively define a class of reverse filtrations with resistances over the set $\Psi^0 = \mathscr{W}$ of irreducible sets for P_0; assume $\mathscr{W}$ has $N_{\mathscr{W}}$ elements, with $N_{\mathscr{W}} \geq 2$. As before, for $\psi, \phi \in \mathscr{W}$ the resistance $r^0(\psi, \phi)$ is the least resistance of any direct path from $x \in \psi$ to the target ϕ with forbidden set $W = \{x\} \cup \Omega \setminus \Omega(x)$. Now starting with Ψ^{k-1} we observe that there is at least one non-trivial circuit, and that every singleton element is trivially a circuit. Hence we can form a non-trivial partition of Ψ^{k-1} into circuits, and denote this partition Ψ^k. As before we define the modified resistance $R^{k-1}(\psi^{k-1}, \phi^{k-1}) = r^{k-1}(\psi^{k-1}, \phi^{k-1}) - r^{k-1}(\psi^{k-1})$, and the resistance function on Ψ^k by the least modified resistance: $r^k(\psi^k, \phi^k) = \min_{\psi^{k-1} \in \psi^k, \phi^{k-1} \in \phi^k} R^{k-1}(\psi^{k-1}, \phi^{k-1})$. Note that since each partition is non-trivial, this construction has at most $k \leq N_{\mathscr{W}}$ layers before the partition has a single element and the construction stops.

Given a reverse filtration, for given $x \in \Omega(x)$ we can now define $\psi^k(x)$ recursively by $x \in \psi^o(x), \psi^0(x) \in \psi^1(x) \ldots$. The *modified radius* of $x \in \Omega(x)$ of order k is then defined by

$$\overline{R}^k(x) = \sum_{\kappa=0}^{k} r^\kappa(\psi^\kappa(x)).$$

Then

Theorem. *[Theorem 9 in the text] Let k be such that $\psi^k(x) = \psi^k(y)$; then $r_x - r_y = \overline{R}^{k-1}(y) - \overline{R}^{k-1}(x)$ and consequently*

$$\frac{C^N}{N^{N-2} D^N} \epsilon^{\overline{R}^{k-1}(y) - \overline{R}^{k-1}(x)} \leq \frac{\mu_\epsilon(x)}{\mu_\epsilon(y)} \leq \frac{N^{N-2} D^N}{C^N} \epsilon^{\overline{R}^{k-1}(y) - \overline{R}^{k-1}(x)}.$$

Proof. From Lemma 14 we know that $r_x - r_y = r^0_{\psi^0(x)} - r^0_{\psi^0(y)}$. Applying Lemma 15 iteratively, we see that if $\psi^{k-1} \in \psi^k$ then

$$r^0_{\psi^0} = r^k_{\psi^k} + \sum_{\kappa=0}^{k-1} \left[\sum_{\phi^\kappa \in \Psi^\kappa} r^\kappa(\phi^\kappa) \right] - \sum_{\kappa=0}^{k-1} r^\kappa(\psi^\kappa)$$

from which

$$r^0_{\psi^0(x)} - r^0_{\psi^0(y)} = -\sum_{\kappa=0}^{k-1} r^\kappa(\psi^\kappa(x)) + \sum_{\kappa=0}^{k-1} r^\kappa(\psi^\kappa(y)) = \overline{R}^{k-1}(y) - \overline{R}^{k-1}(x).$$

□

References

Acemoglu, Daron and James A. Robinson (2001): "A Theory of Political Transitions," *The American Economic Review* **91**: 938-963.

Aircraft Accident Investigation Bureau (2002), "Final Report concerning the accident to the aircraft AVRO 146-RJ100, HB-IXM, operated by Crossair under flight number CRX 3597, on 24 November 2001 near Bassersdorf/ZH," Report 1793.

Ellison, Glenn (2000): "Basins of Attraction, Long Run Stochastic Stability and the Speed of Step-by-step Evolution," *Review of Economic Studies* 67: 17-45.

Fenby, Jonathan (2008): *Modern China: The Fall and Rise of a Great Power, 1850 to the Present,* Penguin Press.

Kandori, Michihiro, George Mailath, and Rafael Rob (1993): "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica* 61: 29-56.

Levine, David and Salvatore Modica (2012): "Conflict, Evolution, Hegemony and the Power of the State," working paper, www.dklevine.com.

Young, Peyton (1993): "The Evolution of Conventions," *Econometrica* 61: 57-84.