Mechanisms for Generating Power-Law Size Distributions, Part 1

Last updated: 2025/10/02, 11:51:22 EDT

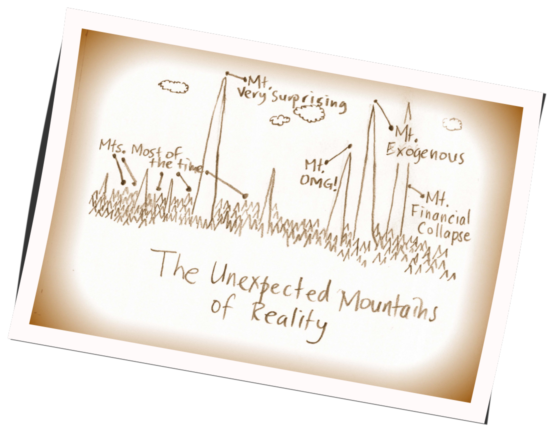
Principles of Complex Systems, Vols. 1, 2, 3D, 4 Fourever, V for Vendetta

Prof. Peter Sheridan Dodds

Computational Story Lab | Vermont Complex Systems Institute
University of Vermont | Santa Fe Institute

Outline

- Heavy-tailed distributions are **characters**.
- Some of these distributions have power-law tails.
- Measured exponents (γ's and α's) vary across systems (and measurers).
- What's their **origin story**?



Mechanisms:

A powerful story in the rise of complexity:

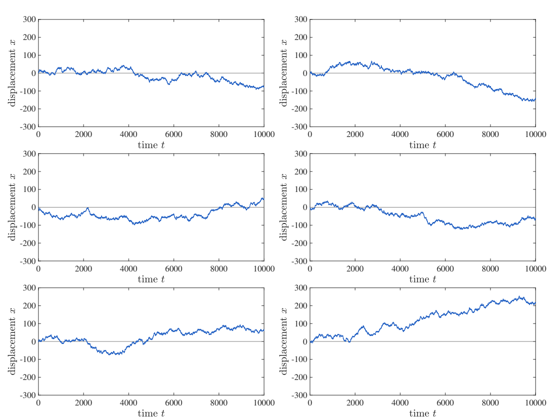
- structure arises out of randomness.
- Exhibit A: Random walks.

The essential random walk:

- One spatial dimension.
- Time and space are discrete
- Random walker (e.g., a zombie texter) starts at origin $x = 0$.
- Step at time t is ε_t:

$$\varepsilon_t = \begin{cases} +1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$$

A few random random walks:



Random walks:

Displacement after t steps:

$$x_t = \sum_{i=1}^{t} \varepsilon_i$$

Expected displacement:

$$\langle x_t \rangle = \left\langle \sum_{i=1}^{t} \varepsilon_i \right\rangle = \sum_{i=1}^{t} \langle \varepsilon_i \rangle = 0$$

- At any time step, we 'expect' our zombie texter to be back at their starting place.
- Obviously fails for odd number of steps...
- But as time goes on, the chance of our texting undead friend lurching back to x=0 must diminish, right?

Variances sum:*

$$\mathrm{Var}(x_t) = \mathrm{Var}\left(\sum_{i=1}^{t} \varepsilon_i\right)$$

$$= \sum_{i=1}^{t} \mathrm{Var}\,(\varepsilon_i) = \sum_{i=1}^{t} 1 = t$$

* Sum rule = a good reason for using the variance to measure spread; only works for independent distributions.

So typical displacement from the origin scales as:

$$\boxed{\sigma = t^{1/2}}$$

- A non-trivial scaling law arises out of additive aggregation or accumulation.

Random walk basics:

Counting random walks:

- Each specific random walk of length t appears with a chance $1/2^t$.
- We'll be more interested in how many random walks end up at the same place.
- Define $N(i, j, t)$ as # distinct walks that start at $x = i$ and end at $x = j$ after t time steps.
- Random walk must displace by $+(j - i)$ after t steps.
- Insert assignment question

$$N(i, j, t) = \binom{t}{(t + j - i)/2}$$

How does $P(x_t)$ behave for large t?

- Take time $t = 2n$ to help ourselves.
- $x_{2n} \in \{0, \pm 2, \pm 4, \ldots, \pm 2n\}$
- x_{2n} is even so set $x_{2n} = 2k$.
- Using our expression $N(i, j, t)$ with $i = 0$, $j = 2k$, and $t = 2n$, we have

$$\mathbf{Pr}(x_{2n} \equiv 2k) \propto \binom{2n}{n + k}$$

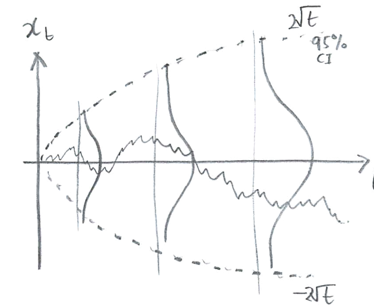
- For large n, the binomial deliciously approaches the Normal Distribution of Snoredom:

$$\mathbf{Pr}(x_t \equiv x) \simeq \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}.$$

Insert assignment question

- The whole is different from the parts. #nutritious
- See also: Stable Distributions

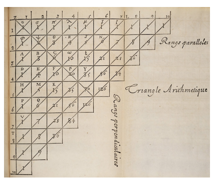
Universality is also not left-handed:



- This is Diffusion: the most essential kind of spreading (more later).
- View as Random Additive Growth Mechanism.

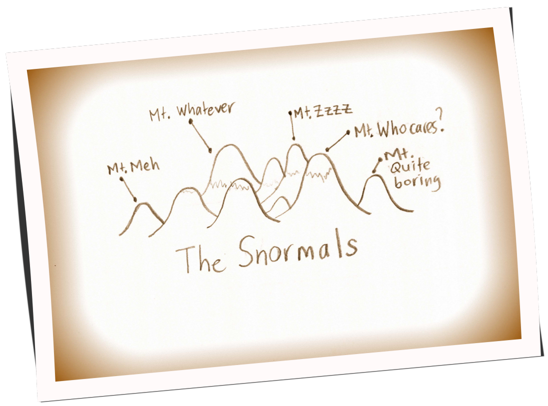
So many things are connected:

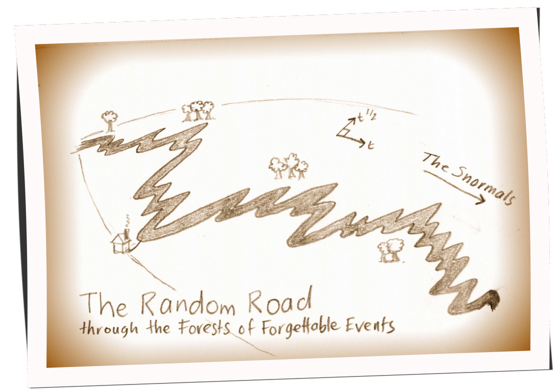
Pascal's Triangle



- Could have been the Pyramid of Pingala[1] or the Triangle of Khayyam, Jia Xian, Tartaglia, ...
- Binomials tend towards the Normal.
- Counting encoded in algebraic forms (and much more).
- Encode heads and tails as variables h and t.
- $(h + t)^n = \sum_{k=0}^{n} \binom{n}{k} h^k t^{n-k}$ where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
- $(h + t)^3 = hhh + hht + hth + thh + htt + tht + tth + ttt$

[1] Stigler's Law of Eponymy showing excellent form again.



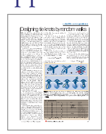


Random walks are even weirder than you might think...

- $\xi_{r,t}$ = the probability that by time step t, a random walk has crossed the origin r times.
- Think of a coin flip game with ten thousand tosses.
- If you are behind early on, what are the chances you will make a comeback?
- The most likely number of lead changes is... 0.
- In fact: $\xi_{0,t} > \xi_{1,t} > \xi_{2,t} > \cdots$
- Even crazier:
  The expected time between tied scores = ∞

See Feller, Intro to Probability Theory, Volume I [5]

Applied knot theory:



"Designing tie knots by random walks"
Fink and Mao,
Nature, **398**, 31–32, 1999. [6]

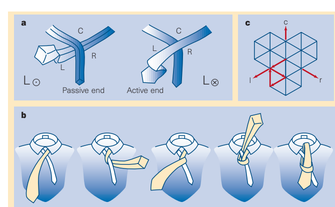


Figure 1 All diagrams are drawn in the frame of reference of the mirror image of the actual tie. **a,** The two ways of beginning a knot, $L_{\odot}$ and $L_{\otimes}$. For knots beginning with $L_{\odot}$, the tie must begin inside-out. **b,** The four-in-hand, denoted by the sequence $L_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$. **c,** A knot may be represented by a persistent random walk on a triangular lattice. The example shown is the four-in-hand, indicated by the walk ↑↑↑ĉ.

Hexagons are the bestagons.

Applied knot theory:

Table 1 **Aesthetic tie knots**

h	γ	γ/h	$K(h,\gamma)$	s	b	Name	Sequence
3	1	0.33	1	0	0		$L_{\odot}R_{\otimes}C_{\odot}T$
4	1	0.25	1	−1	1	Four-in-hand	$L_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$
5	2	0.40	2	−1	0	Pratt knot	$L_{\odot}C_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$
6	2	0.33	4	0	0	Half-Windsor	$L_{\otimes}R_{\odot}C_{\otimes}L_{\odot}R_{\otimes}C_{\odot}T$
7	2	0.29	6	−1	1		$L_{\odot}R_{\otimes}L_{\odot}C_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$
7	3	0.43	4	0	1		$L_{\odot}C_{\otimes}R_{\odot}C_{\otimes}L_{\odot}R_{\otimes}C_{\odot}T$
8	2	0.25	8	0	2		$L_{\otimes}R_{\odot}L_{\otimes}C_{\odot}R_{\otimes}L_{\odot}R_{\otimes}C_{\odot}T$
8	3	0.38	12	−1	0	Windsor	$L_{\otimes}C_{\odot}R_{\otimes}L_{\odot}C_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$
9	3	0.33	24	0	0		$L_{\odot}R_{\otimes}C_{\odot}L_{\otimes}R_{\odot}C_{\otimes}L_{\odot}R_{\otimes}C_{\odot}T$
9	4	0.44	8	−1	2		$L_{\odot}C_{\otimes}R_{\odot}C_{\otimes}L_{\odot}C_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$

Knots are characterized by half-winding number h, centre number γ, centre fraction γ/h, knots per class $K(h,\gamma)$, symmetry s, balance b, name and sequence.

- h = number of moves
- γ = number of center moves
- $K(h, \gamma) = 2^{\gamma-1}\binom{h-\gamma-2}{\gamma-1}$
- $s = \sum_{i=1}^{h} x_i$ where $x_i = -1$ for L and $x_i = +1$ for R.
- $b = \frac{1}{2}\sum_{i=2}^{h-1} |\omega_i + \omega_{i-1}|$ where $\omega = \pm 1$ represents winding direction.

Random walks

#crazytownbananapants

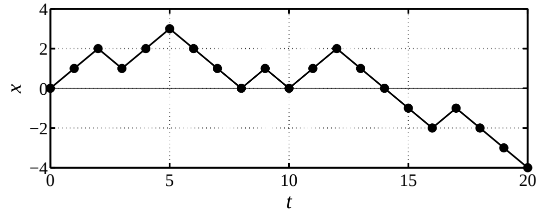
The problem of first return:

- What is the probability that a random walker in one dimension returns to the origin for the first time after t steps?
- Will our zombie texter always return to the origin?
- What about higher dimensions?

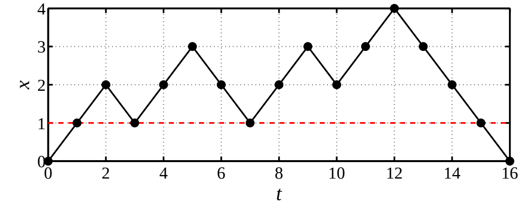
Reasons for caring:

1. We will find a power-law size distribution with an interesting exponent.
2. Some physical structures may result from random walks.
3. We'll start to see how different scalings relate to each other.

For random walks in 1-d:



- A return to origin can only happen when $t = 2n$.
- In example above, returns occur at $t = 8$, 10, and 14.
- Call $P_{\rm fr}(2n)$ the probability of first return at $t = 2n$.
- Probability calculation ≡ Counting problem (combinatorics/statistical mechanics).
- Idea: Transform first return problem into an easier return problem.



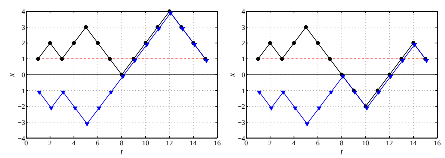
- Can assume zombie texter first lurches to $x = 1$.
- Observe walk first returning at $t = 16$ stays at or above $x = 1$ for $1 \le t \le 15$ (dashed red line).
- Now want walks that can return many times to $x = 1$.
- $P_{\rm fr}(2n) =$
  $2 \cdot \frac{1}{2} Pr(x_t \ge 1, 1 \le t \le 2n-1, \text{ and } x_1 = x_{2n-1} = 1)$
- The $\frac{1}{2}$ accounts for $x_{2n} = 2$ instead of 0.
- The 2 accounts for texters that first lurch to $x = -1$.

Counting first returns:

Approach:

- Move to counting numbers of walks.
- Return to probability at end.
- Again, $N(i, j, t)$ is the # of possible walks between $x = i$ and $x = j$ taking t steps.
- Consider all paths starting at $x = 1$ and ending at $x = 1$ after $t = 2n - 2$ steps.
- Idea: If we can compute the number of walks that hit $x = 0$ at least once, then we can subtract this from the total number to find the ones that maintain $x \ge 1$.
- Call walks that drop below $x = 1$ excluded walks.
- We'll use a method of images to identify these excluded walks.

Examples of excluded walks:



Key observation for excluded walks:

- For any path starting at x=1 that hits 0, there is a unique matching path starting at x=−1.
- Matching path first mirrors and then tracks after first reaching x=0.
- # of t-step paths starting and ending at x=1 and hitting x=0 at least once
  = # of t-step paths starting at x=−1 and ending at x=+1
  = $N(-1, +1, t)$

- Call the number of paths that return after $t = 2n$ time steps after first moving to the positive side $N^+_{\rm fr}(2n)$.
- For paths that first move to the negative side: $N^-_{\rm fr}(2n)$.
- So $N^+_{\rm fr}(2n) = N(+1, +1, 2n-2) - N(-1, +1, 2n-2)$
- Negative side:
  $N^-_{\rm fr}(2n) = N(-1, -1, 2n-2) - N(+1, -1, 2n-2)$
- Symmetry: $N^+_{\rm fr}(2n) = N^-_{\rm fr}(2n)$
- Both $N_{\rm fr}(2n)$ and the one sided $N^+_{\rm fr}(2n)$ are of mathematical and physical interest.
- Overall:

$$N_{\rm fr}(2n) = N^+_{\rm fr}(2n) + N^-_{\rm fr}(2n) = 2N^+_{\rm fr}(2n)$$

$$= 2N(+1, +1, 2n-2) - 2N(-1, +1, 2n-2).$$

One of many related things: Catalan numbers

Probability of first return:

Insert assignment question:

- Find

$$N_{\rm fr}(2n) \sim \frac{2^{2n-3/2}}{\sqrt{2\pi} n^{3/2}}.$$

- Normalized number of paths gives probability.
- Total number of possible paths $= 2^{2n}$.
-

$$P_{\rm fr}(2n) = \frac{1}{2^{2n}} N_{\rm fr}(2n)$$

$$\simeq \frac{1}{2^{2n}} \frac{2^{2n-3/2}}{\sqrt{2\pi} n^{3/2}}$$

$$= \frac{1}{\sqrt{2\pi}} (2n)^{-3/2} \propto t^{-3/2}.$$

- We have $P(t) \propto t^{-3/2}$, $\gamma = 3/2$.
- Same scaling holds for continuous space/time walks.
- $P(t)$ is normalizable.
- Recurrence: Random walker always returns to origin
- But mean, variance, and all higher moments are infinite. #totalmadness
- Even though walker must return, expect a long wait...
- One moral: Repeated gambling against an infinitely wealthy opponent must lead to ruin.

Higher dimensions:

- Walker in $d = 2$ dimensions must also return
- Walker may not return in $d \ge 3$ dimensions
- Associated human ~~genius~~: George Pólya

Random walks

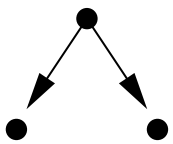
On finite spaces:

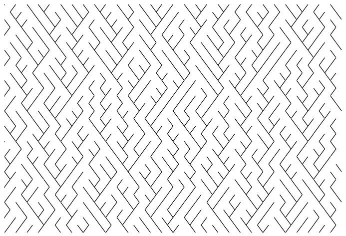
- In any finite homogeneous space, a random walker will visit every site with equal probability
- Call this probability the Invariant Density of a dynamical system
- Non-trivial Invariant Densities arise in chaotic systems.

On networks:

- On networks, a random walker visits each node with frequency ∝ node degree #groovy
- Equal probability still present: walkers traverse edges with equal frequency. #totallygroovy

Scheidegger Networks [17, 4]





- Random directed network on triangular lattice.
- Toy model of real networks.
- 'Flow' is southeast or southwest with equal probability.

Scheidegger networks

- Creates basins with random walk boundaries.
- Observe that subtracting one random walk from another gives random walk with increments:

$$\varepsilon_t = \begin{cases} +1 & \text{with probability } 1/4 \\ 0 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/4 \end{cases}$$

- Random walk with probabilistic pauses.
- Basin termination = first return random walk problem.
- Basin length ℓ distribution: $P(\ell) \propto \ell^{-3/2}$
- For real river networks, generalize to $P(\ell) \propto \ell^{-\gamma}$.

Connections between exponents:

- For a basin of length ℓ, width $\propto \ell^{1/2}$
- Basin area $a \propto \ell \cdot \ell^{1/2} = \ell^{3/2}$
- Invert: $\ell \propto a^{2/3}$
- $\mathrm{d}\ell \propto \mathrm{d}(a^{2/3}) = 2/3 a^{-1/3} \mathrm{d}a$
- $\mathbf{Pr}(\text{basin area} = a)\mathrm{d}a$
  $= \mathbf{Pr}(\text{basin length} = \ell)\mathrm{d}\ell$
  $\propto \ell^{-3/2}\mathrm{d}\ell$
  $\propto (a^{2/3})^{-3/2} a^{-1/3}\mathrm{d}a$
  $= a^{-4/3}\mathrm{d}a$
  $= a^{-\tau}\mathrm{d}a$

Connections between exponents:

- Both basin area and length obey power law distributions
- Observed for real river networks
- Reportedly: $1.3 < \tau < 1.5$ and $1.5 < \gamma < 2$

Generalize relationship between area and length:

- Hack's law [10]:

$$\ell \propto a^h.$$

- For real, large networks [13] $h \simeq 0.5$ (isometric scaling)
- Smaller basins possibly $h > 1/2$ (allometric scaling).
- Models exist with interesting values of h.
- Plan: Redo calc with γ, τ, and h.

Connections between exponents:

- Given

$$\ell \propto a^h, \; P(a) \propto a^{-\tau}, \text{ and } P(\ell) \propto \ell^{-\gamma}$$

- $\mathrm{d}\ell \propto \mathrm{d}(a^h) = h a^{h-1}\mathrm{d}a$
- Find τ in terms of γ and h.
- $\mathbf{Pr}(\text{basin area} = a)\mathrm{d}a$
  $= \mathbf{Pr}(\text{basin length} = \ell)\mathrm{d}\ell$
  $\propto \ell^{-\gamma}\mathrm{d}\ell$
  $\propto (a^h)^{-\gamma} a^{h-1}\mathrm{d}a$
  $= a^{-(1+h(\gamma-1))}\mathrm{d}a$
- $$\boxed{\tau = 1 + h(\gamma - 1)}$$
- Excellent example of the Scaling Relations found between exponents describing power laws for many systems.

Connections between exponents:

With more detailed description of network structure, $\tau = 1 + h(\gamma - 1)$ simplifies to: [3]

$$\boxed{\tau = 2 - h}$$

and

$$\boxed{\gamma = 1/h}$$

- Only one exponent is independent (take h).
- Simplifies system description.
- Expect Scaling Relations where power laws are found.
- Need only characterize Universality class with independent exponents.

Death ...

Failure:

- A very simple model of failure/death
- x_t = entity's 'health' at time t
- Start with $x_0 > 0$.
- Entity fails when x hits 0.

"Explaining mortality rate plateaus"
Weitz and Fraser,
Proc. Natl. Acad. Sci., **98**, 15383–15386, 2001. [18]

... and the NBA:

Basketball and other sports [2]:

- Three arcsine laws (Lévy [12]) for continuous-time random walk lasting time T:

$$\frac{1}{\pi}\frac{1}{\sqrt{t(T-t)}}.$$

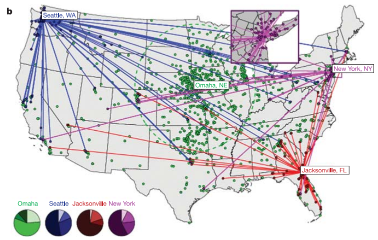
  The arcsine distribution applies for:
  (1) fraction of time positive,
  (2) the last time the walk changes sign,
  and (3) the time the maximum is achieved.
- Well approximated by basketball score lines [8, 2].
- Australian Rules Football has some differences [11].

More than randomness

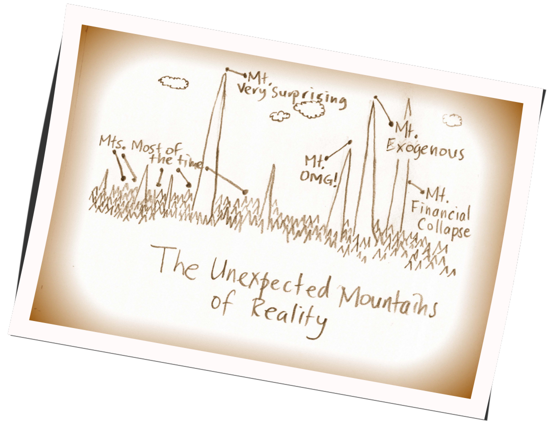
- Can generalize to Fractional Random Walks [15, 16, 14]
- Fractional Brownian Motion, Lévy flights
- See Montroll and Shlesinger for example: [14]
  "On $1/f$ noise and other distributions with long tails."
  Proc. Natl. Acad. Sci., 1982.
- In 1-d, standard deviation σ scales as

$$\sigma \sim t^{\alpha}$$

  $\alpha = 1/2$ — diffusive
  $\alpha > 1/2$ — superdiffusive
  $\alpha < 1/2$ — subdiffusive
- Extensive memory of path now matters...



- First big studies of movement and interactions of people.
- Brockmann *et al.* [1] "Where's George" study.
- Beyond Lévy: Superdiffusive in space but with long waiting times.
- Tracking movement via cell phones [9] and Twitter [7].



References I

[1] D. Brockmann, L. Hufnagel, and T. Geisel.
The scaling laws of human travel.
Nature, pages 462–465, 2006. pdf

[2] A. Clauset, M. Kogan, and S. Redner.
Safe leads and lead changes in competitive team sports.
Phys. Rev. E, 91:062815, 2015. pdf

[3] P. S. Dodds and D. H. Rothman.
Unified view of scaling laws for river networks.
Physical Review E, 59(5):4865–4877, 1999. pdf

[4] P. S. Dodds and D. H. Rothman.
Scaling, universality, and geomorphology.
Annu. Rev. Earth Planet. Sci., 28:571–610, 2000. pdf

References II

[5] W. Feller.
An Introduction to Probability Theory and Its Applications, volume I.
John Wiley & Sons, New York, third edition, 1968.

[6] T. M. Fink and Y. Mao.
Designing tie knots by random walks.
Nature, 398:31–32, 1999. pdf

[7] M. R. Frank, L. Mitchell, P. S. Dodds, and C. M. Danforth.
Happiness and the patterns of life: A study of geolocated Tweets.
Nature Scientific Reports, 3:2625, 2013. pdf

[8] A. Gabel and S. Redner.
Random walk picture of basketball scoring.
Journal of Quantitative Analysis in Sports, 8:1–20, 2012.

References III

[9] M. C. González, C. A. Hidalgo, and A.-L. Barabási.
Understanding individual human mobility patterns.
Nature, 453:779–782, 2008. pdf

[10] J. T. Hack.
Studies of longitudinal stream profiles in Virginia and Maryland.
United States Geological Survey Professional Paper, 294-B:45–97, 1957. pdf

[11] D. P. Kiley, A. J. Reagan, L. Mitchell, C. M. Danforth, and P. S. Dodds.
The game story space of professional sports: Australian Rules Football.
Physical Review E, 93, 2016. pdf

References IV

[12] P. Lévy and M. Loeve.
Processus stochastiques et mouvement brownien.
Gauthier-Villars Paris, 1965.

[13] D. R. Montgomery and W. E. Dietrich.
Channel initiation and the problem of landscape scale.
Science, 255:826–30, 1992. pdf

[14] E. W. Montroll and M. F. Shlesinger.
On the wonderful world of random walks, volume XI of Studies in statistical mechanics, chapter 1, pages 1–121.
New-Holland, New York, 1984.

[15] E. W. Montroll and M. W. Shlesinger.
On $1/f$ noise and other distributions with long tails.
Proc. Natl. Acad. Sci., 79:3380–3383, 1982. pdf

References V

[16] E. W. Montroll and M. W. Shlesinger.
Maximum entropy formalism, fractals, scaling phenomena, and $1/f$ noise: a tale of tails.
J. Stat. Phys., 32:209–230, 1983.

[17] A. E. Scheidegger.
The algebra of stream-order numbers.
United States Geological Survey Professional Paper, 525-B:B187–B189, 1967. pdf

[18] J. S. Weitz and H. B. Fraser.
Explaining mortality rate plateaus.
Proc. Natl. Acad. Sci., 98:15383–15386, 2001. pdf