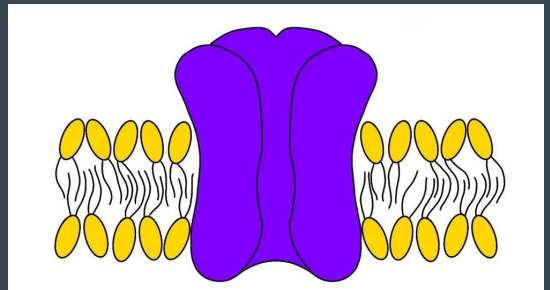
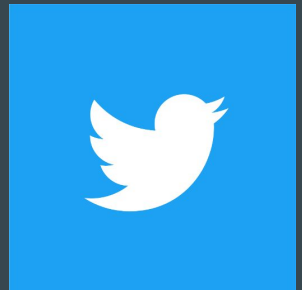
# Baseball's Unprecedented Half-innings and Other Insights from Markov Chains

...

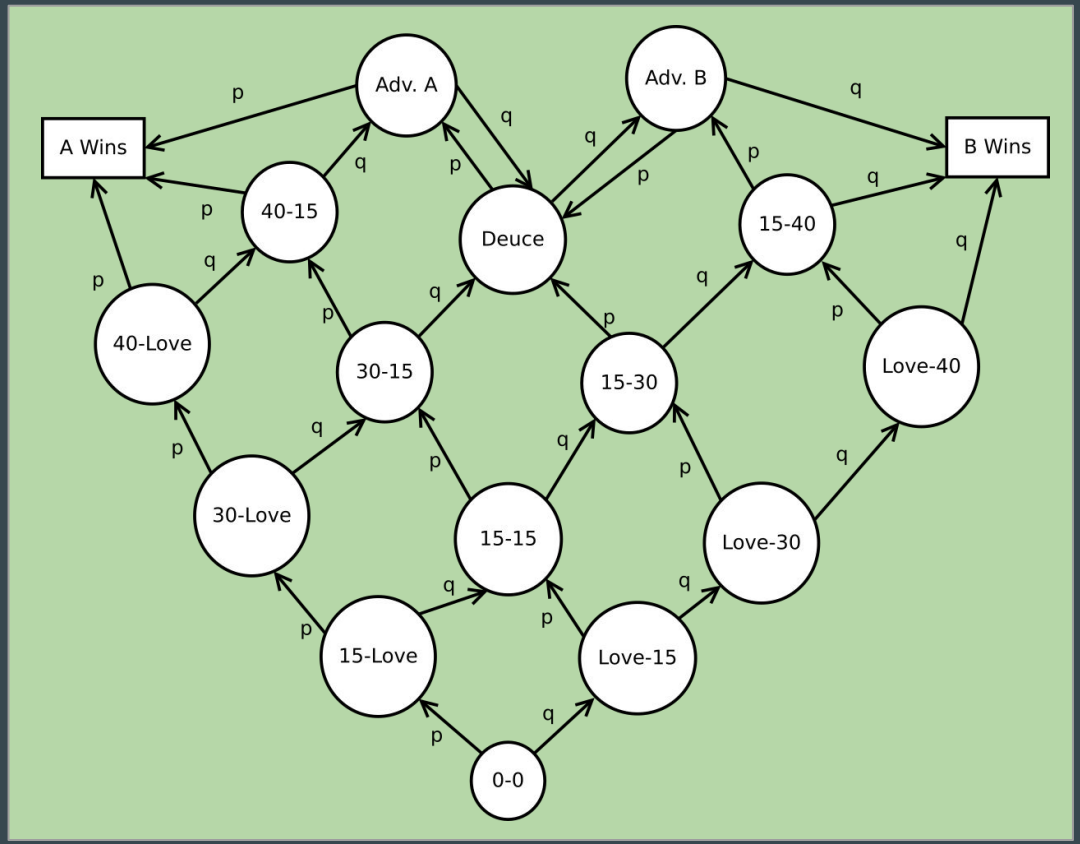
Sean G. Carver, Ph.D.

September 12, 2019

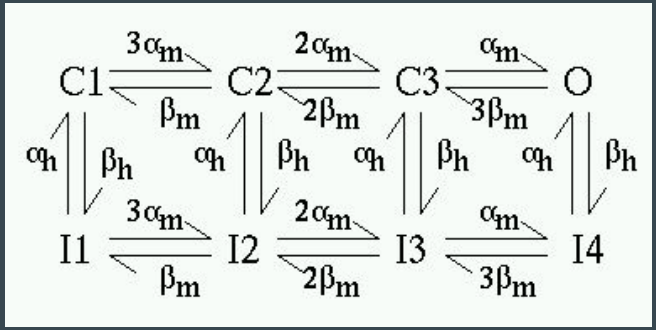




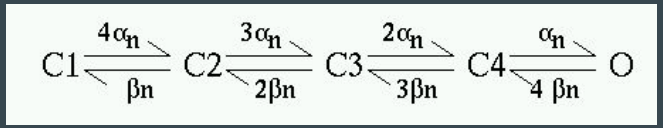
# What is a Markov Chain?



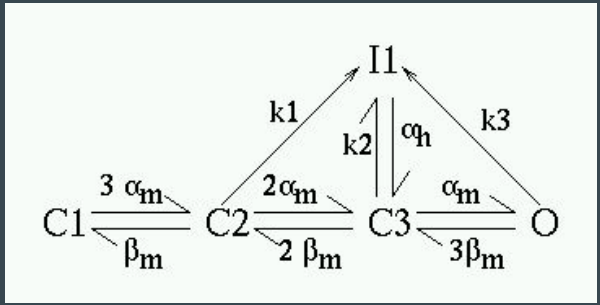
# Markov Chains Beyond Sports




Hodgkin-Huxley Sodium Channel




Potassium Channel

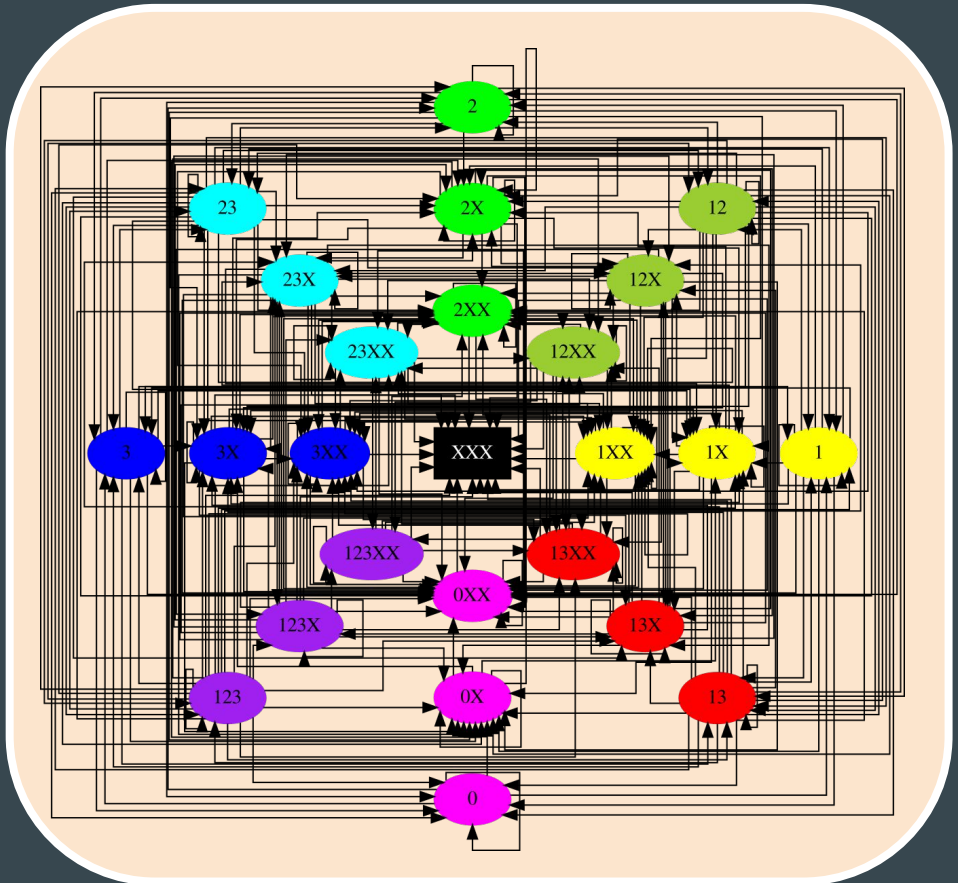



Modern Model of Sodium Channel

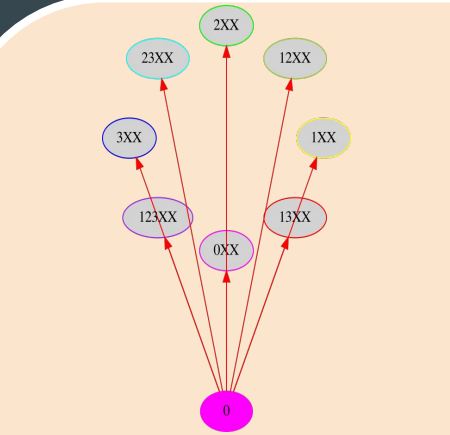
Exposure
Incubation
Transmission
Immunity

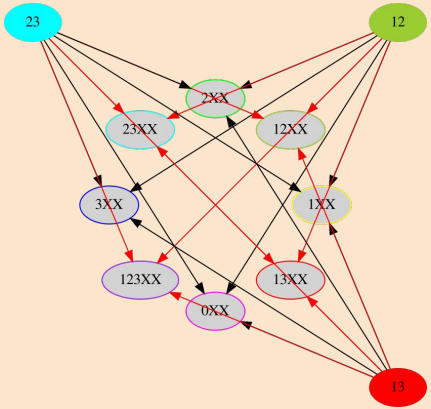
Detect statistically significant changes to users' or communities' interactions with Twitter

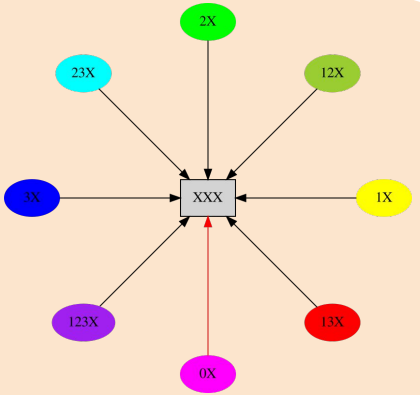
# Baseball as a Markov Chain

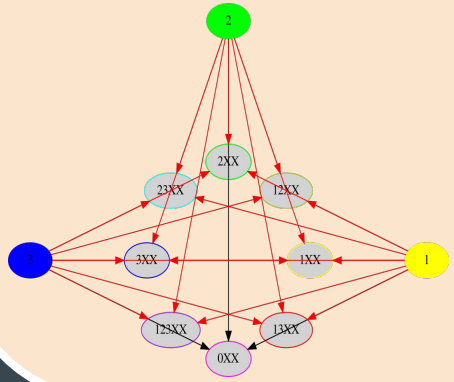



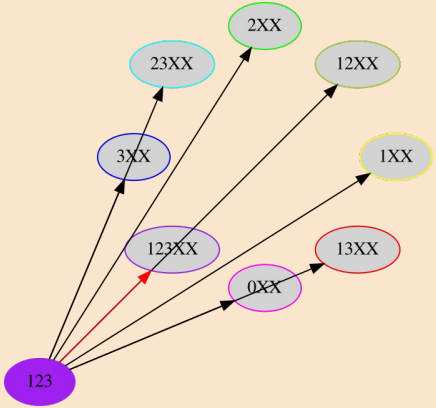
# Visualizing Baseball's Graph (Double and Triple Plays)

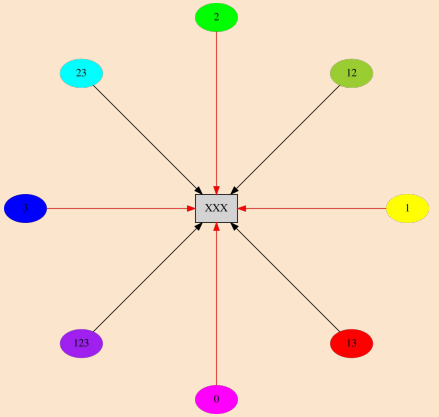


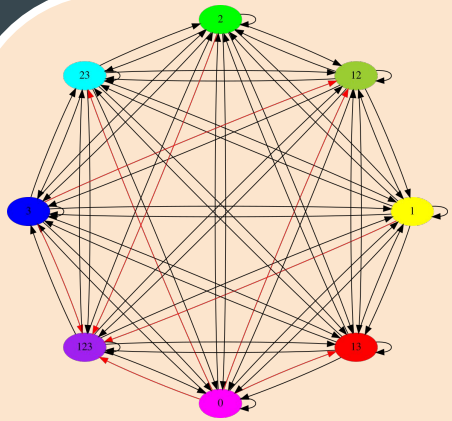



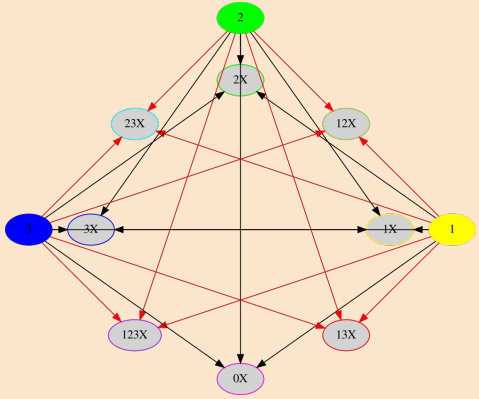



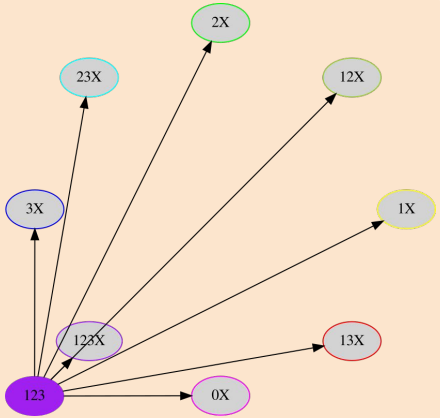



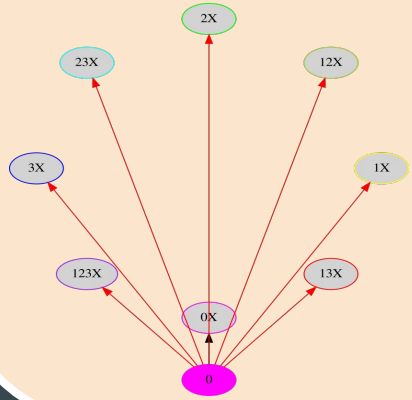



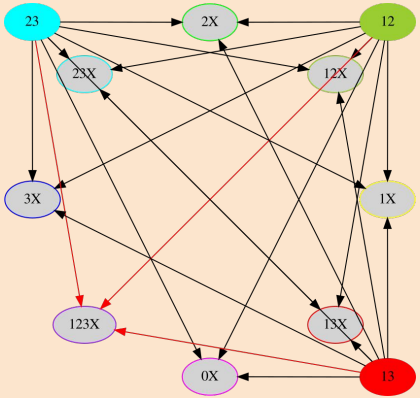
# Visualizing Baseball's Graph (Clean and Single Plays)

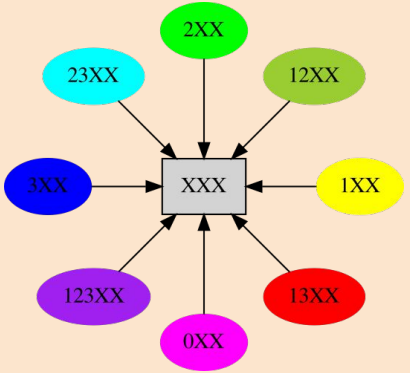


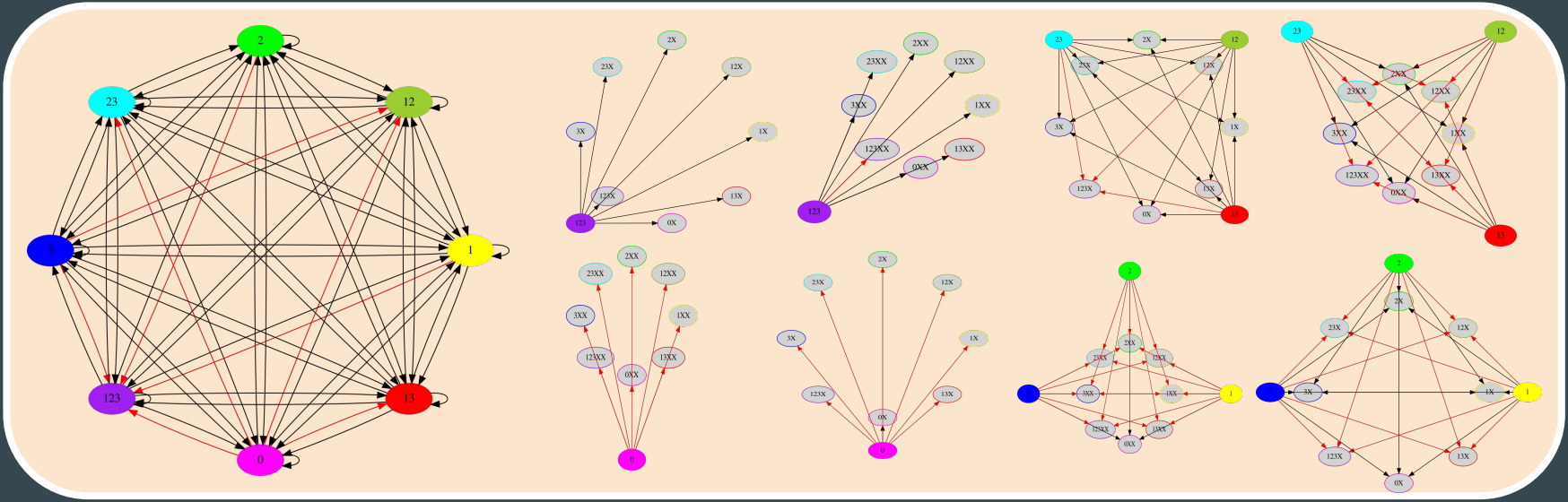



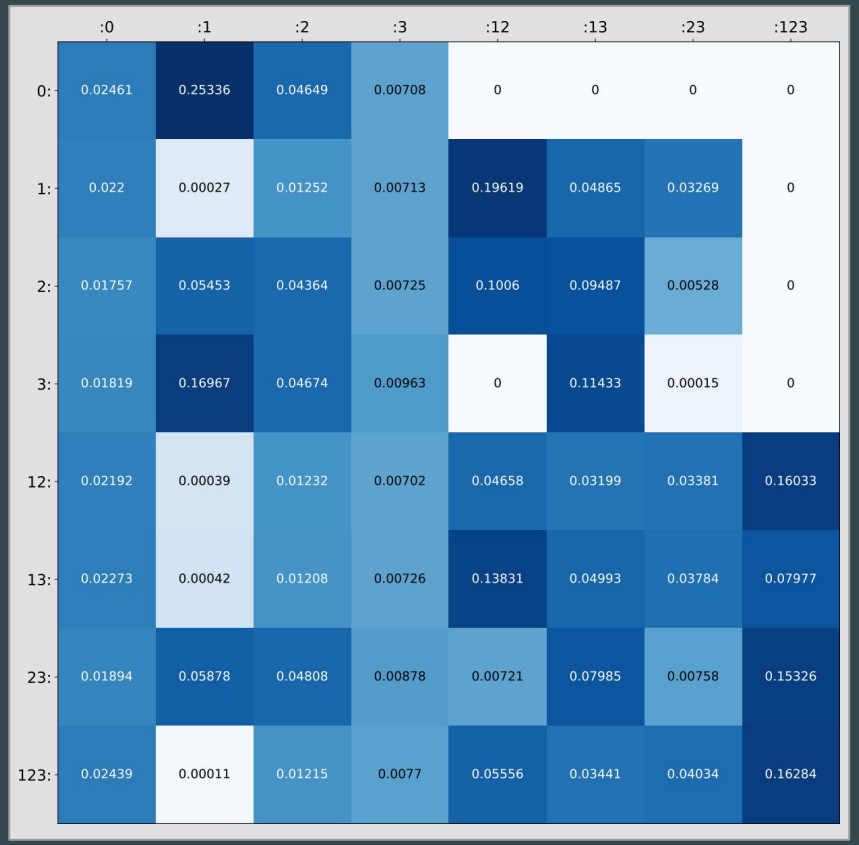
# Symmetry and Asymmetry in Baseball's Graph




Rules of symmetric & real baseball: Δ Outs ≥ 0 AND 0 ≤ Runners-on-base ≤ 3
Δ Runners-on-base = 1 - Δ Outs - Δ Score
Δ Score ≥ 0

Additional asymmetries in real baseball: Three outs and it's over!
3:12 (3X:12X, ...) impossible---runners cannot retreat!

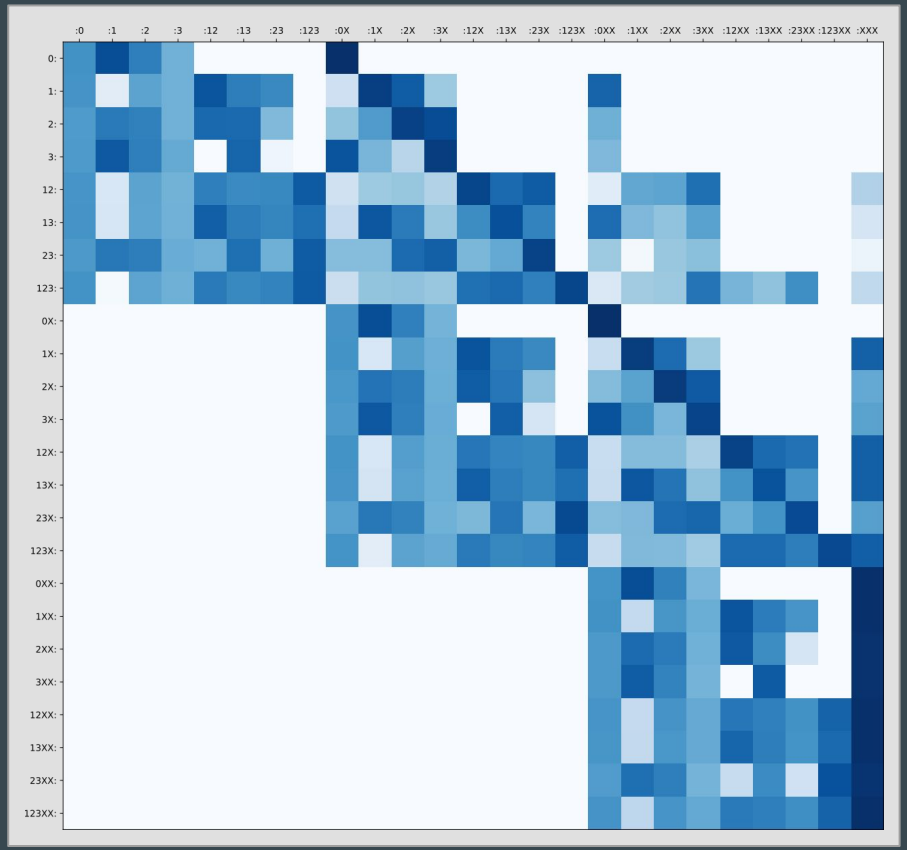
# Transition Probability Matrix -- Clean Plays, No Outs




$$\text{Prob.} = \frac{\#\,(\text{From} \rightarrow \text{To})}{\#\,\text{From}}$$

Computed with respect to a population of transitions

# Full Transition Probability Matrix

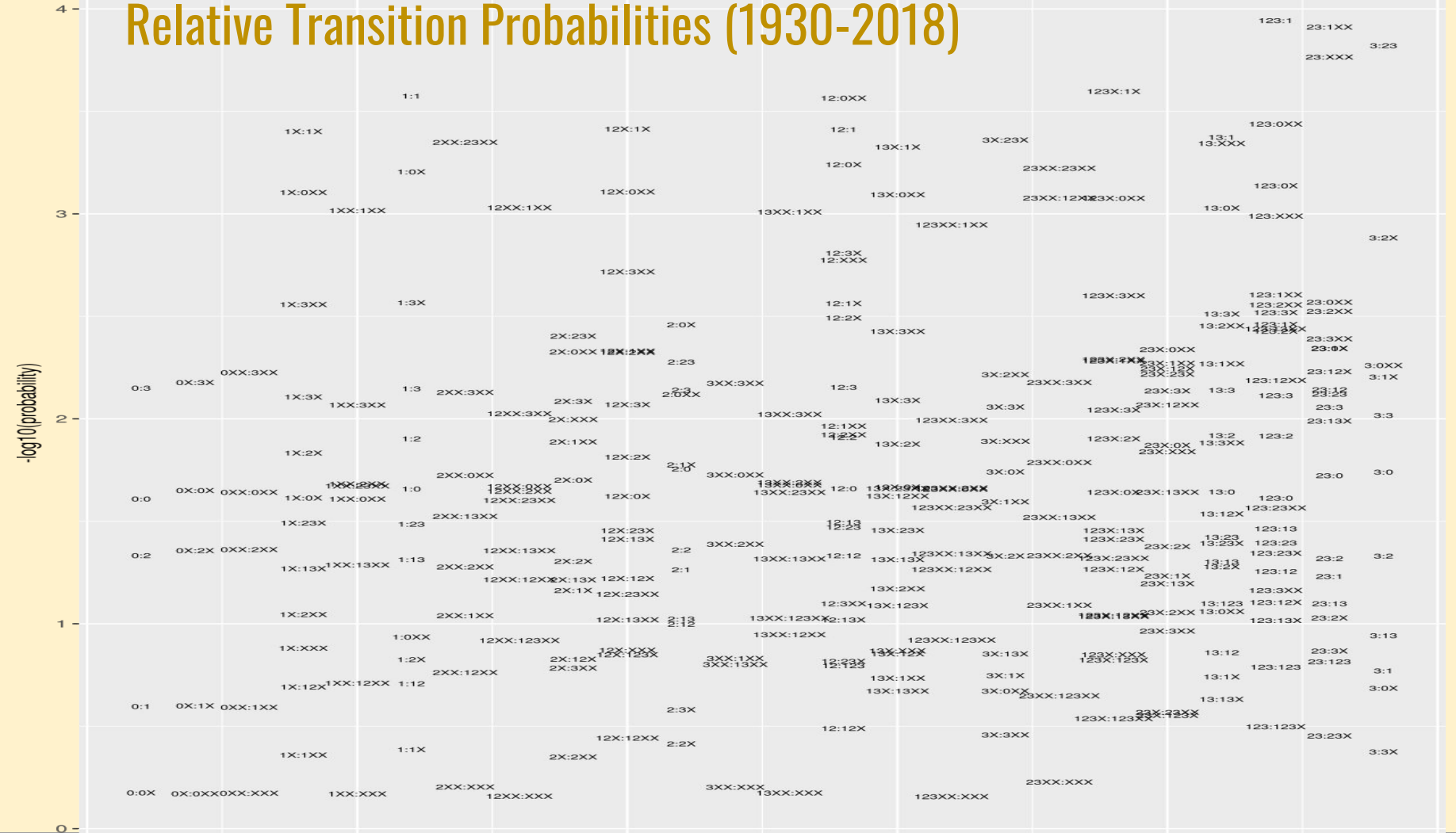



Population:
All MLB teams, 1930-2018,
Regular season

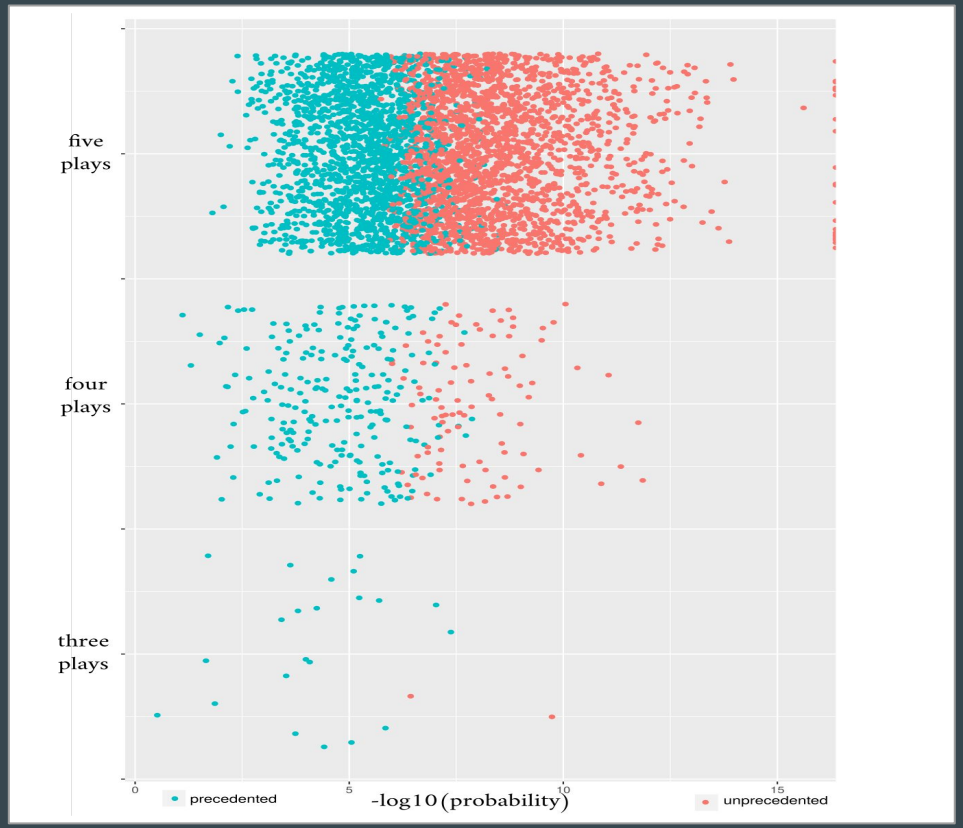
# Many More Probabilities Are Zero for Subpopulations


Washington Nationals, batting at home, 2018

# Relative Transition Probabilities (1930-2018)



# Baseball's Unprecedented Half-innings

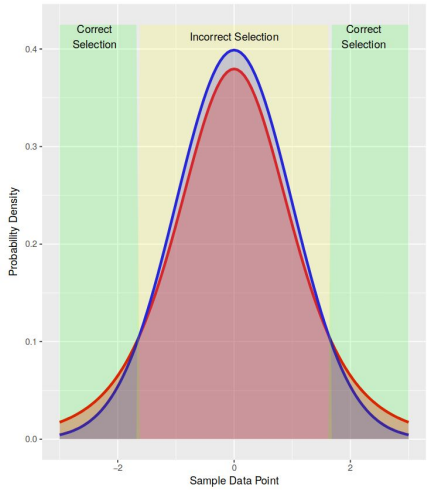


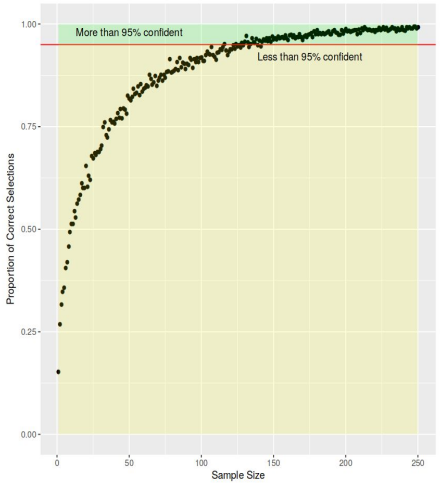

The most unlikely unprecedented half-innings:

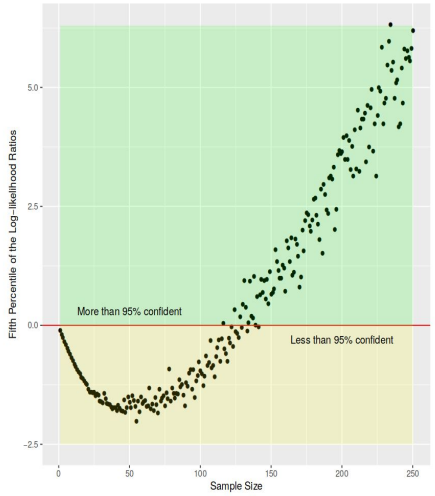
| sequence | -log10(probability) |
|---|---|
| Three Plays | |
| 0:3:13:XXX | 6.43509767499117 |
| 0:3:23:XXX | 9.73543779430501 |
| Four Plays | |
| 0:2:3:23:XXX | 11.0579993507313 |
| 0:0:3:23:XXX | 11.3442887847059 |
| 0:3:3:23:XXX | 11.7519898723546 |
| 0:3:23:23:XXX | 11.855734799529 |
| Five Plays | |
| 0:3:23:3:23:XXX | 15.6122744428876 |
| *3XX:23XX* | Infinity |

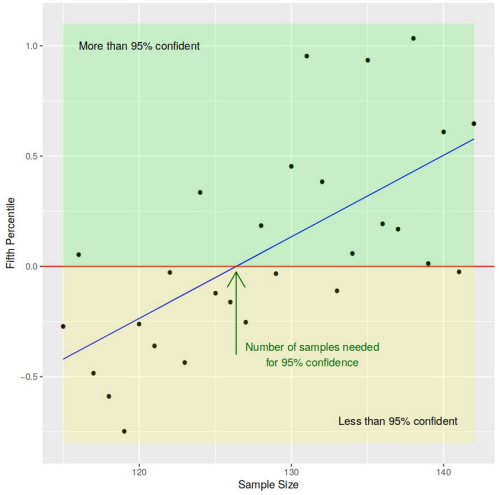
log10(number of half-innings played) = 6.45

# Quantifying the Similarity Between Model Baseball Teams




?


Result: it takes 30±1 half-innings, simulated from the 2011 Baltimore Orioles model, to reject, with 95% confidence, the statement that these half-innings were sampled from the 2011 New York Yankees model.

# Making Half-inning History and Markov Transition Probabilities Easily Accessible to All

# Future work: Applying Markov Chains to Win Games

The Markov model can be used to detect statistically significant changes, including...

- Changes to a team's play,
- Changes to the whole league's play.

Once a change has been detected, insight into why can be pursued.

This insight can be leveraged to win games.

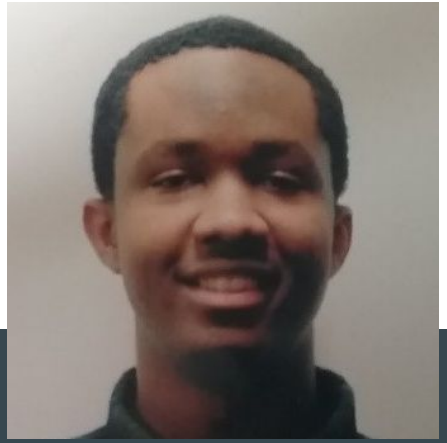
# Collaborators on Related Projects

## Jacob Ward

Formerly Professorial Lecturer, American University

## Jake Berberian

Class of 2022, American University



## Kingsley Iyawe

Masters Student, American University, expected graduation: May 2020