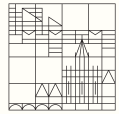
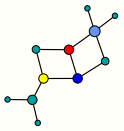


Social shifts in the Late Pre–hispanic US Southwest

Habiba, Jan C. Athenstädt & Ulrik Brandes

Department of Computer & Information Science
University of Konstanz

April 17, 2015



Agenda

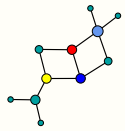
Introduction

- History
- Mills et al. (2013):Transformation of social networks in the late pre–Hispanic US Southwest
- B-R: Under the magnifying glass
- Follow–up questions

Beyond Brainerd–Robinson

- Alternative measures of similarity
- Across–Time Comparison

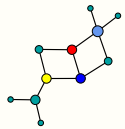
ViSim - A tool to explore similarities among sites



Introduction

History

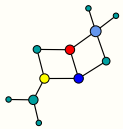
- US Southwest (A.D. 1200–1450): large-scale demographic changes
  - long-distance migration (from north to south in late 1200s)
  - population aggregation (in south in 1300s)



Introduction

History

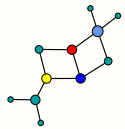
- US Southwest (A.D. 1200–1450): large-scale demographic changes
  - long-distance migration (from north to south in late 1200s)
  - population aggregation (in south in 1300s)



Introduction

History

- US Southwest (A.D. 1200–1450): large-scale demographic changes
  - long-distance migration (from north to south in late 1200s)
  - population aggregation (in south in 1300s)



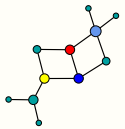
Introduction

Mills et al. (2013): Transformation of social networks in the late pre–Hispanic US Southwest

- Reconstruct population dynamics using network approach
- Database: 42 distinct artifacts, 700+ sites/settlements, over 250 years
  - 515 settlements with ≥ 30 artifacts
  - discretized 250 years into 50-years periods
- Similarity: Brainerd-Robinson index

$$BR(x, y) = 200 - \sum_{z=1}^{p} |P_{xz} - P_{yz}|$$

- Assumption: exchanges movement migrations emulations = Similarity in consumption of wares



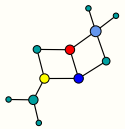
Introduction

Mills et al. (2013): Transformation of social networks in the late pre–Hispanic US Southwest

- Reconstruct population dynamics using network approach
- Database: 42 distinct artifacts, 700+ sites/settlements, over 250 years
  - 515 settlements with ≥ 30 artifacts
  - discretized 250 years into 50-years periods
- Similarity: Brainerd-Robinson index

$$BR(x, y) = 200 - \sum_{z=1}^{p} |P_{xz} - P_{yz}|$$

- Assumption: exchanges movement migrations emulations = Similarity in consumption of wares



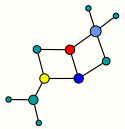
Introduction

Mills et al. (2013): Transformation of social networks in the late pre–Hispanic US Southwest

- Reconstruct population dynamics using network approach
- Database: 42 distinct artifacts, 700+ sites/settlements, over 250 years
  - 515 settlements with ≥ 30 artifacts
  - discretized 250 years into 50-years periods
- Similarity: Brainerd-Robinson index

$$BR(x, y) = 200 - \sum_{z=1}^{p} |P_{xz} - P_{yz}|$$

- Assumption: exchanges movement migrations emulations = Similarity in consumption of wares



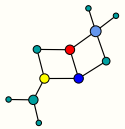
Introduction

Mills et al. (2013): Transformation of social networks in the late pre–Hispanic US Southwest

- Reconstruct population dynamics using network approach
- Database: 42 distinct artifacts, 700+ sites/settlements, over 250 years
  - 515 settlements with ≥ 30 artifacts
  - discretized 250 years into 50-years periods
- Similarity: Brainerd-Robinson index

$$BR(x, y) = 200 - \sum_{z=1}^{p} |P_{xz} - P_{yz}|$$

- Assumption: exchanges movement migrations emulations = Similarity in consumption of wares



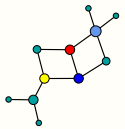
Introduction

Mills et al. (2013): Transformation of social networks in the late pre–Hispanic US Southwest

- Reconstruct population dynamics using network approach
- Database: 42 distinct artifacts, 700+ sites/settlements, over 250 years
  - 515 settlements with ≥ 30 artifacts
  - discretized 250 years into 50-years periods
- Similarity: Brainerd-Robinson index

$$BR(x, y) = 200 - \sum_{z=1}^{p} |P_{xz} - P_{yz}|$$

- Assumption: exchanges movement migrations emulations = Similarity in consumption of wares



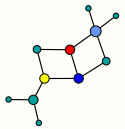
Introduction

Mills et al. (2013): Transformation of social networks in the late pre–Hispanic US Southwest

- Reconstruct population dynamics using network approach
- Database: 42 distinct artifacts, 700+ sites/settlements, over 250 years
  - 515 settlements with ≥ 30 artifacts
  - discretized 250 years into 50-years periods
- Similarity: Brainerd-Robinson index

$$BR(x, y) = 200 - \sum_{z=1}^{p} |P_{xz} - P_{yz}|$$

- Assumption: exchanges movement migrations emulations = Similarity in consumption of wares



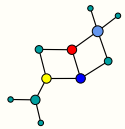
Introduction

Mills et al. (2013): Transformation of social networks in the late pre–Hispanic US Southwest

- Reconstruct population dynamics using network approach
- Database: 42 distinct artifacts, 700+ sites/settlements, over 250 years
  - 515 settlements with ≥ 30 artifacts
  - discretized 250 years into 50-years periods
- Similarity: Brainerd-Robinson index

$$BR(x, y) = 200 - \sum_{z=1}^{p} |P_{xz} - P_{yz}|$$

- Assumption: exchanges movement migrations emulations = Similarity in consumption of wares

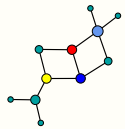


B-R: Under the magnifying glass

Sampling bias

	Site A	Site B	Site C
Type 1	80%	80%	0%
Type 2	15%	5%	75%
Type 3	5%	15%	25%

- $BR(A, B) = 180 \quad BR(A, C) = 40$
- $BR(A, C) = 100$

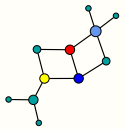


B-R: Under the magnifying glass

Sampling bias

	Site A	Site B	Site C
Type 1	80%	80%	0%
Type 2	15%	5%	75%
Type 3	5%	15%	25%

- $BR(A, B) = 180 \quad BR(A, C) = 40$
- $BR(A, C) = 100$

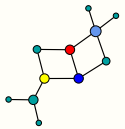


B-R: Under the magnifying glass

Sampling bias

	Site A	Site B	Site C
Type 1	80%	80%	0%
Type 2	15%	5%	75%
Type 3	5%	15%	25%

- $BR(A, B) = 180 \quad BR(A, C) = 40$
- $BR(A, C) = 100$

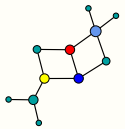


B-R: Under the magnifying glass

Sampling bias

	Site A	Site B	Site C
Type 1	80%	80%	0%
Type 2	15%	5%	75%
Type 3	5%	15%	25%

- $BR(A, B) = 180 \quad BR(A, C) = 40$
- $BR(A, C) = 100$

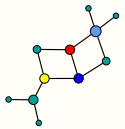


B-R: Under the magnifying glass

Sampling bias

	Site A	Site B	Site C
Type 1	80%	80%	0%
Type 2	15%	5%	75%
Type 3	5%	15%	25%

- $BR(A, B) = 180 \quad BR(A, C) = 40$
- $BR(A, C) = 100$

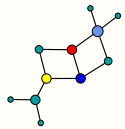


B-R: Under the magnifying glass

Symmetric similarity

	Site A	Site B	Site C
Type 1	40%	20%	20%
Type 2	30%	80%	0%
Type 3	30%	0%	80%

- $BR(A, B) = 100 \quad BR(A, C) = 100 \quad BR(B, C) = 40$
- $B, C \subset A$

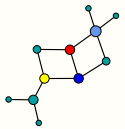


B-R: Under the magnifying glass

Symmetric similarity

	Site A	Site B	Site C
Type 1	40%	20%	20%
Type 2	30%	80%	0%
Type 3	30%	0%	80%

- $BR(A, B) = 100 \quad BR(A, C) = 100 \quad BR(B, C) = 40$
- $B, C \subset A$



B-R: Under the magnifying glass

Symmetric similarity

	Site A	Site B	Site C
Type 1	40%	20%	20%
Type 2	30%	80%	0%
Type 3	30%	0%	80%

- $BR(A, B) = 100 \quad BR(A, C) = 100 \quad BR(B, C) = 40$
- $B, C \subset A$

B-R: Under the magnifying glass

Symmetric similarity

	Site A	Site B	Site C
Type 1	40%	20%	20%
Type 2	30%	80%	0%
Type 3	30%	0%	80%

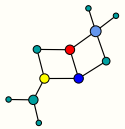
- $BR(A, B) = 100 \quad BR(A, C) = 100 \quad BR(B, C) = 40$
- $B, C \subset A$

B-R: Under the magnifying glass

Aggregation

	Site A	Site B	Site C	Site D
Type 1	80%	100%	25%	35%
Type 2	5%	0%	35%	40%
Type 3	5%	0%	30%	20%
Type 4	10%	0%	10%	5%

- $BR(A, B) = 160 \quad BR(C, D) = 160$
- C, D both contain all types and differ by atmost 10 % in quantity
- A, B, B contains only Type 1 whereas A contains all the types.



B-R: Under the magnifying glass

Aggregation

	Site A	Site B	Site C	Site D
Type 1	80%	100%	25%	35%
Type 2	5%	0%	35%	40%
Type 3	5%	0%	30%	20%
Type 4	10%	0%	10%	5%

- $BR(A, B) = 160 \quad BR(C, D) = 160$
- C, D both contain all types and differ by atmost 10 % in quantity
- A, B, B contains only Type 1 whereas A contains all the types.

B-R: Under the magnifying glass

Aggregation

	Site A	Site B	Site C	Site D
Type 1	80%	100%	25%	35%
Type 2	5%	0%	35%	40%
Type 3	5%	0%	30%	20%
Type 4	10%	0%	10%	5%

- $BR(A, B) = 160 \quad BR(C, D) = 160$
- C, D both contain all types and differ by atmost 10 % in quantity
- A, B, B contains only Type 1 whereas A contains all the types.

B-R: Under the magnifying glass

Aggregation

	Site A	Site B	Site C	Site D
Type 1	80%	100%	25%	35%
Type 2	5%	0%	35%	40%
Type 3	5%	0%	30%	20%
Type 4	10%	0%	10%	5%

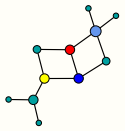
- $BR(A, B) = 160 \quad BR(C, D) = 160$
- C, D both contain all types and differ by atmost 10 % in quantity
- A, B, B contains only Type 1 whereas A contains all the types.

B-R: Under the magnifying glass

Aggregation

	Site A	Site B	Site C	Site D
Type 1	80%	100%	25%	35%
Type 2	5%	0%	35%	40%
Type 3	5%	0%	30%	20%
Type 4	10%	0%	10%	5%

- $BR(A, B) = 160 \quad BR(C, D) = 160$
- C, D both contain all types and differ by atmost 10 % in quantity
- A, B, B contains only Type 1 whereas A contains all the types.

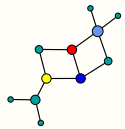


B-R: Under the magnifying glass

Aggregation

	Site A	Site B	Site C	Site D
Type 1	80%	100%	25%	35%
Type 2	5%	0%	35%	40%
Type 3	5%	0%	30%	20%
Type 4	10%	0%	10%	5%

- $BR(A, B) = 160 \quad BR(C, D) = 160$
- C, D both contain all types and differ by atmost 10 % in quantity
- A, B, B contains only Type 1 whereas A contains all the types.



Introduction

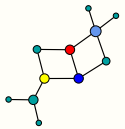
Follow-up questions

1. How do larger and more diverse settlements relate to the smaller and more homogeneous ones?
2. How does population shifts happened within shorter or longer time periods?
3. How much is the evolving “identity” of settlements indicative of movement trends?

Introduction

Follow-up questions

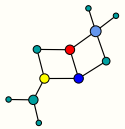
1. How do larger and more diverse settlements relate to the smaller and more homogeneous ones?
2. How does population shifts happened within shorter or longer time periods?
3. How much is the evolving “identity” of settlements indicative of movement trends?



Introduction

Follow-up questions

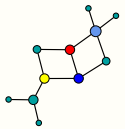
1. How do larger and more diverse settlements relate to the smaller and more homogeneous ones?
2. How does population shifts happened within shorter or longer time periods?
3. How much is the evolving “identity” of settlements indicative of movement trends?



Introduction

Follow-up questions

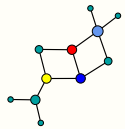
1. How do larger and more diverse settlements relate to the smaller and more homogeneous ones?
2. How does population shifts happened within shorter or longer time periods?
3. How much is the evolving “identity” of settlements indicative of movement trends?



Reconstructing social shifts

Proposed Extensions

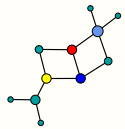
1. Asymmetric similarity - dominance relationship
2. Ranking of wares/types
3. Index of significance of wares/types
4. Across–time comparison



Reconstructing social shifts

Proposed Extensions

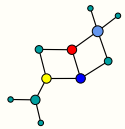
1. Asymmetric similarity - dominance relationship
2. Ranking of wares/types
3. Index of significance of wares/types
4. Across–time comparison



Reconstructing social shifts

Proposed Extensions

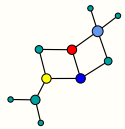
1. Asymmetric similarity - dominance relationship
2. Ranking of wares/types
3. Index of significance of wares/types
4. Across–time comparison



Reconstructing social shifts

Proposed Extensions

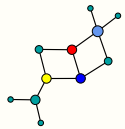
1. Asymmetric similarity - dominance relationship
2. Ranking of wares/types
3. Index of significance of wares/types
4. Across–time comparison



Reconstructing social shifts

Proposed Extensions

1. Asymmetric similarity - dominance relationship
2. Ranking of wares/types
3. Index of significance of wares/types
4. Across–time comparison



Reconstructing social shifts

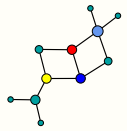
Asymmetric similarity based on dominance relationship

- Integral: A site x is dominated by a site y or the site x is completely contained in the site y if and only if the set of distinct items found on site x is a proper subset of the set of distinct items found on site y.

$$S_R(x,y) = \begin{cases} 1 & \text{if } B_{x,i} \geq B_{y,i} \forall i \in [1,n] \\ 0 & \text{otherwise} \end{cases}$$

- Fractional: A site x is dominated by another site y, if each type present in x is also present in y. It is strictly dominated, if it is dominated and there is at least one type in y that is not present in x.

$$S_R(x,y) = \begin{cases} 1 & \text{if } Q_{x,i} > Q_{y,i} \forall i \in [1,n] \\ 0 & \text{otherwise} \end{cases}$$



Reconstructing social shifts

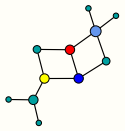
Asymmetric similarity based on dominance relationship

- Integral: A site x is dominated by a site y or the site x is completely contained in the site y if and only if the set of distinct items found on site x is a proper subset of the set of distinct items found on site y.

$$S_R(x,y) = \begin{cases} 1 & \text{if } B_{x,i} \geq B_{y,i} \forall i \in [1,n] \\ 0 & \text{otherwise} \end{cases}$$

- Fractional: A site x is dominated by another site y, if each type present in x is also present in y. It is strictly dominated, if it is dominated and there is at least one type in y that is not present in x.

$$S_R(x,y) = \begin{cases} 1 & \text{if } Q_{x,i} > Q_{y,i} \forall i \in [1,n] \\ 0 & \text{otherwise} \end{cases}$$



Reconstructing social shifts

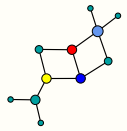
Asymmetric similarity based on dominance relationship

- Integral: A site x is dominated by a site y or the site x is completely contained in the site y if and only if the set of distinct items found on site x is a proper subset of the set of distinct items found on site y.

$$S_R(x,y) = \begin{cases} 1 & \textit{if } \; B_{x,i} \geq B_{y,i} \forall i \in [1,n] \\ 0 & \textit{otherwise} \end{cases}$$

- Fractional: A site x is dominated by another site y, if each type present in x is also present in y. It is strictly dominated, if it is dominated and there is at least one type in y that is not present in x.

$$S_R(x,y) = \begin{cases} 1 & \textit{if } \; Q_{x,i} > Q_{y,i} \forall i \in [1,n] \\ 0 & \textit{otherwise} \end{cases}$$



Reconstructing social shifts

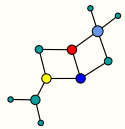
Similarity based on relative ranking of wares

- Parametrized: k-out-of-top-l
  A pair of sites are similar to each other if they have l of k top ranked wares common among them.

$$S_R(x,y) = \begin{cases} 1 & \text{if } |V_R^x[1:k] \cap V_R^y[1:k]| \geq l \\ 0 & \text{otherwise} \end{cases}$$

- Non–parametrized: Maximum Quasi–Jaccard
  A pair of sites are k similar to each other for the maximal k of p types that they can be similar in.

$$S_R(x,y) = \arg\max_k \frac{|V_R^x[1:k] \cap V_R^y[1:k]|}{|V_R^x[1:k] \cup V_R^y[1:k]|}$$



Reconstructing social shifts

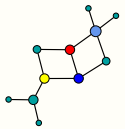
Similarity based on relative ranking of wares

- Parametrized: k-out-of-top-l
  A pair of sites are similar to each other if they have l of k top ranked wares common among them.

$$S_R(x,y) = \begin{cases} 1 & \text{if } |V_R^x[1:k] \cap V_R^y[1:k]| \geq l \\ 0 & \text{otherwise} \end{cases}$$

- Non–parametrized: Maximum Quasi–Jaccard
  A pair of sites are k similar to each other for the maximal k of p types that they can be similar in.

$$S_R(x,y) = \arg\max_k \frac{|V_R^x[1:k] \cap V_R^y[1:k]|}{|V_R^x[1:k] \cup V_R^y[1:k]|}$$



Reconstructing social shifts

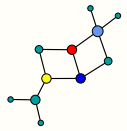
Similarity based on relative ranking of wares

- Parametrized: k-out-of-top-l
  A pair of sites are similar to each other if they have l of k top ranked wares common among them.

$$S_R(x,y) = \begin{cases} 1 & \text{if } |V_R^x[1:k] \cap V_R^y[1:k]| \geq l \\ 0 & \text{otherwise} \end{cases}$$

- Non–parametrized: Maximum Quasi–Jaccard
  A pair of sites are k similar to each other for the maximal k of p types that they can be similar in.

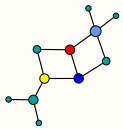
$$S_R(x,y) = \arg\max_k \frac{|V_R^x[1:k] \cap V_R^y[1:k]|}{|V_R^x[1:k] \cup V_R^y[1:k]|}$$



Reconstructing social shifts

Index of Significance of Wares

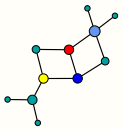
- *TF − IDF*: (*term frequency–inverse document frequency*), is a numerical statistic that is intended to depict the importance of a word in a document.
  - $f(i, x)$: frequency of each ware i in site x.
  - $\frac{|S|}{1+|x \in S : i \in x|}$: inverse the frequency of i in all sites.
  - $I(i, x) = f(i, x) \times \frac{|S|}{1+|x \in S : i \in x|}$
- Similarity among sites based on $I(w_i, x)$
- Co–occurrence of wares
- Evolving “identity” of settlements over periods of time.



Reconstructing social shifts

Index of Significance of Wares

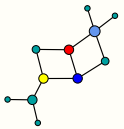
- *TF* − *IDF*: (*term frequency–inverse document frequency*), is a numerical statistic that is intended to depict the importance of a word in a document.
  - $f(i, x)$: frequency of each ware i in site x.
  - $\frac{|S|}{1+|x \in S : i \in x|}$: inverse the frequency of i in all sites.
  - $I(i, x) = f(i, x) \times \frac{|S|}{1+|x \in S : i \in x|}$
- Similarity among sites based on $I(w_i, x)$
- Co–occurrence of wares
- Evolving “identity” of settlements over periods of time.



Reconstructing social shifts

Index of Significance of Wares

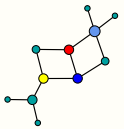
- *TF* − *IDF*: (*term frequency–inverse document frequency*), is a numerical statistic that is intended to depict the importance of a word in a document.
  - $f(i, x)$: frequency of each ware i in site x.
  - $\frac{|S|}{1+|x \in S : i \in x|}$: inverse the frequency of i in all sites.
  - $I(i, x) = f(i, x) \times \frac{|S|}{1+|x \in S : i \in x|}$
- Similarity among sites based on $I(w_i, x)$
- Co–occurrence of wares
- Evolving “identity” of settlements over periods of time.



Reconstructing social shifts

Index of Significance of Wares

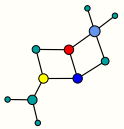
- $TF - IDF$: (*term frequency–inverse document frequency*), is a numerical statistic that is intended to depict the importance of a word in a document.
  - $f(i, x)$: frequency of each ware i in site x.
  - $\frac{|S|}{1+|x \in S : i \in x|}$: inverse the frequency of i in all sites.
  - $I(i, x) = f(i, x) \times \frac{|S|}{1+|x \in S : i \in x|}$
- Similarity among sites based on $I(w_i, x)$
- Co–occurrence of wares
- Evolving “identity” of settlements over periods of time.



Reconstructing social shifts

Index of Significance of Wares

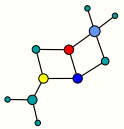
- *TF* − *IDF*: (*term frequency–inverse document frequency*), is a numerical statistic that is intended to depict the importance of a word in a document.
  - $f(i, x)$: frequency of each ware i in site x.
  - $\frac{|S|}{1+|x \in S : i \in x|}$: inverse the frequency of i in all sites.
  - $I(i, x) = f(i, x) \times \frac{|S|}{1+|x \in S : i \in x|}$
- Similarity among sites based on $I(w_i, x)$
- Co–occurrence of wares
- Evolving “identity” of settlements over periods of time.



Reconstructing social shifts

Index of Significance of Wares

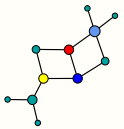
- *TF* − *IDF*: (*term frequency–inverse document frequency*), is a numerical statistic that is intended to depict the importance of a word in a document.
  - $f(i, x)$: frequency of each ware i in site x.
  - $\frac{|S|}{1+|x \in S : i \in x|}$: inverse the frequency of i in all sites.
  - $I(i, x) = f(i, x) \times \frac{|S|}{1+|x \in S : i \in x|}$
- Similarity among sites based on $I(w_i, x)$
- Co–occurrence of wares
- Evolving “identity” of settlements over periods of time.



Reconstructing social shifts

Index of Significance of Wares

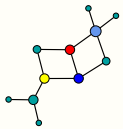
- $TF - IDF$: (*term frequency–inverse document frequency*), is a numerical statistic that is intended to depict the importance of a word in a document.
  - $f(i, x)$: frequency of each ware i in site x.
  - $\frac{|S|}{1+|x \in S : i \in x|}$: inverse the frequency of i in all sites.
  - $I(i, x) = f(i, x) \times \frac{|S|}{1+|x \in S : i \in x|}$
- Similarity among sites based on $I(w_i, x)$
- Co–occurrence of wares
- Evolving “identity” of settlements over periods of time.



Reconstructing social shifts

Index of Significance of Wares

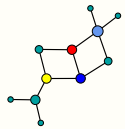
- $TF - IDF$: (*term frequency–inverse document frequency*), is a numerical statistic that is intended to depict the importance of a word in a document.
  - $f(i, x)$: frequency of each ware i in site x.
  - $\frac{|S|}{1+|x \in S : i \in x|}$: inverse the frequency of i in all sites.
  - $I(i, x) = f(i, x) \times \frac{|S|}{1+|x \in S : i \in x|}$
- Similarity among sites based on $I(w_i, x)$
- Co–occurrence of wares
- Evolving “identity” of settlements over periods of time.



Beyond Brainerd–Robinson

Across–Time Comparison

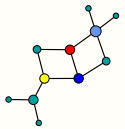
- Long distance movement/migration/resettlement
- Shorter/longer distance movements



Beyond Brainerd–Robinson

Across–Time Comparison

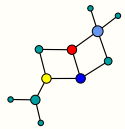
- Long distance movement/migration/resettlement
- Shorter/longer distance movements



Beyond Brainerd–Robinson

Across–Time Comparison

- Long distance movement/migration/resettlement
- Shorter/longer distance movements



ViSim

A tool to explore similarities among settlements