



# ***Genealogical Implicit Affinity Networks***

Matthew Smith and Christophe Giraud-Carrier  
Department of Computer Science  
Brigham Young University

FHTW, March 2006

[smitty@byu.edu](mailto:smitty@byu.edu)



# Outline

- **Introduction**

- Motivation
- Objective

- **Affinity Network Creation**

- **Examples and Discoveries**

- Star Wars (Fictitious)
- My Family (Real)

- **Conclusion**

# Introduction - Motivation



- Evidence suggests that we...
  - Don't know family members as well as we could
  - Forget about them (particularly ancestors)
  - Routinely miss opportunities to become closer
- Plato observed that “similarity begets friendship”
- Discovering what we have in common, i.e., our affinities, with our relatives (both dead and alive) would increase our sense of belonging, allow us to draw strength from others, become more united, and build stronger family ties.



# Objective

- This presentation describes a **method for building networks that highlight affinities**, or inherent similarities, among people, particularly family members.
- The content of such affinity networks can be exploited to **strengthen living families** and to **direct family history research**.
- Preliminary results demonstrate promise.



# Outline

- **Introduction**

- Motivation
- Objective

- **Affinity Network Creation**

- **Examples and Discoveries**

- Star Wars (Fictitious)
- My Family (Real)

- **Conclusion**

# Affinity Network Creation

- Let  $A = \{A_1, A_2, \dots, A_n\}$  (Set of Attributes)
- An individual  $x$  is represented by a tuple  $x = \langle A_1 : a_1^x, A_2 : a_2^x, \dots, A_n : a_n^x \rangle$



# Affinity Network Creation

Individual	Characteristics
Sarah	A B C D E
Bob	A D Q R S
Jim	X Y D
Mary	X Y Z
Susan	R P Q S
Brent	Q

*Metrics generally depend on the nature of the attribute (e.g., nominal, real, string)*

### Common Similarity Metrics

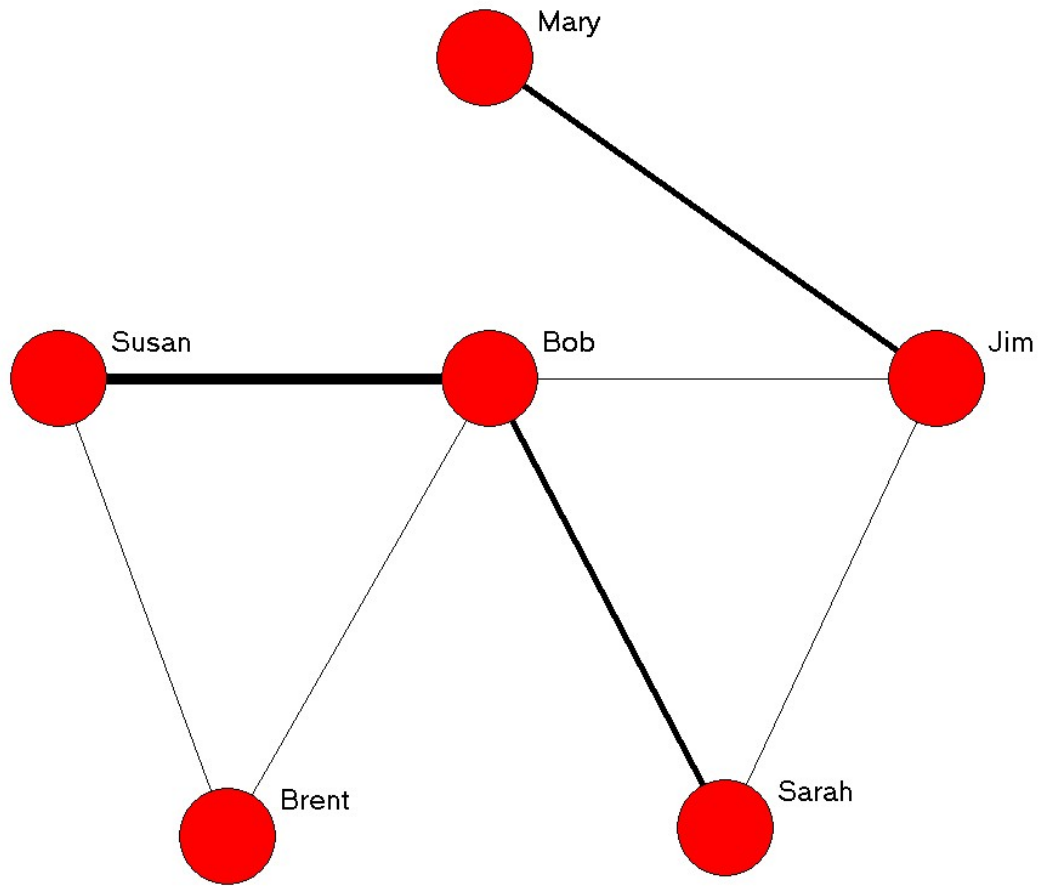
Exact match, Euclidean distance, soundex, metaphone, levenstein, jaro-winkler, jaccard, stemming, etc.

Table 1: Sample of Individuals and their Characteristics

	Sarah	Bob	Jim	Mary	Susan	Brent
Sarah	—	2	1	0	0	0
Bob	2	—	1	0	3	1
Jim	1	1	—	2	0	0
Mary	0	0	2	—	0	0
Susan	0	3	0	0	—	1
Brent	0	1	0	0	1	—

Table 2: Total Affinity Matrix for Individuals in Table 1

# Affinity Network





# Outline

- **Introduction**

- Motivation
- Objective

- **Affinity Network Creation**

- **Examples and Discoveries**

- Star Wars (Fictitious)
- Family Data (Real)

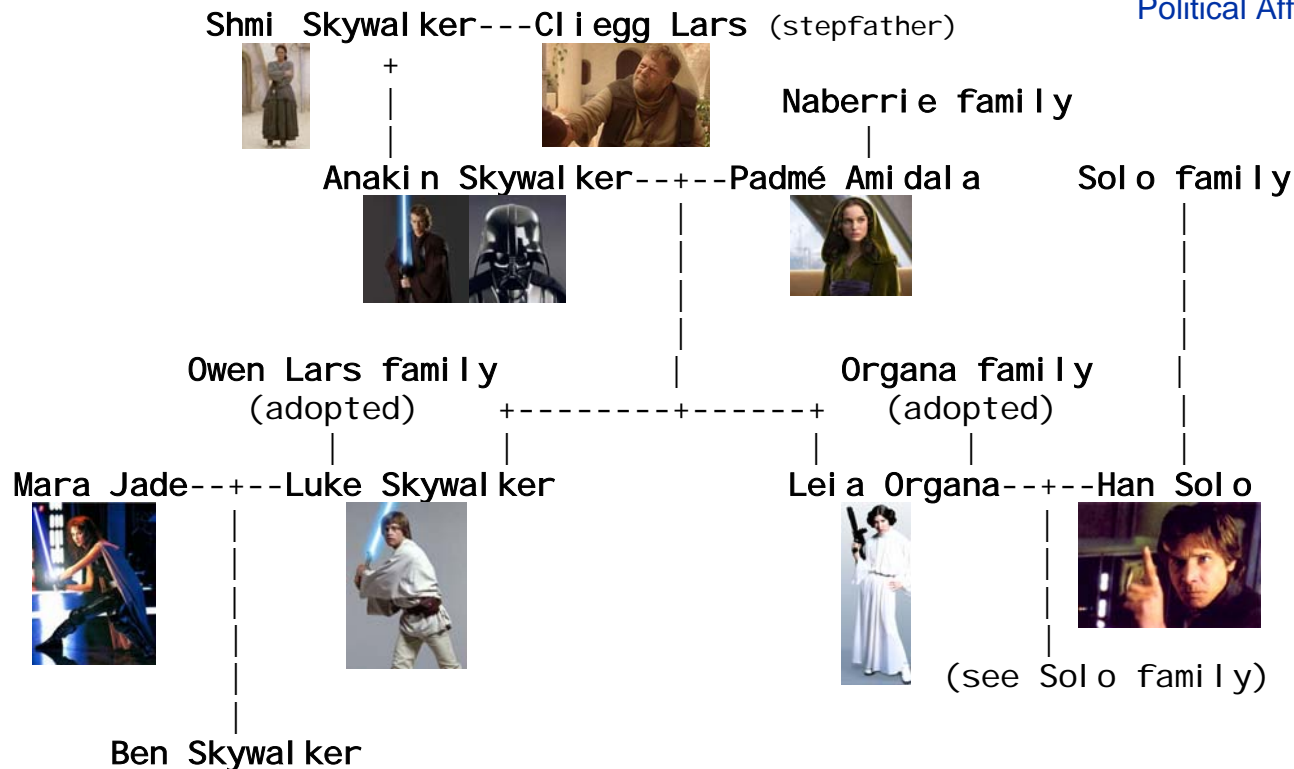
- **Conclusion**



# Skywalker Family Tree

## CHARACTERISTICS

Name, Sex, Hometown, Occupation, Political Affiliation, Children

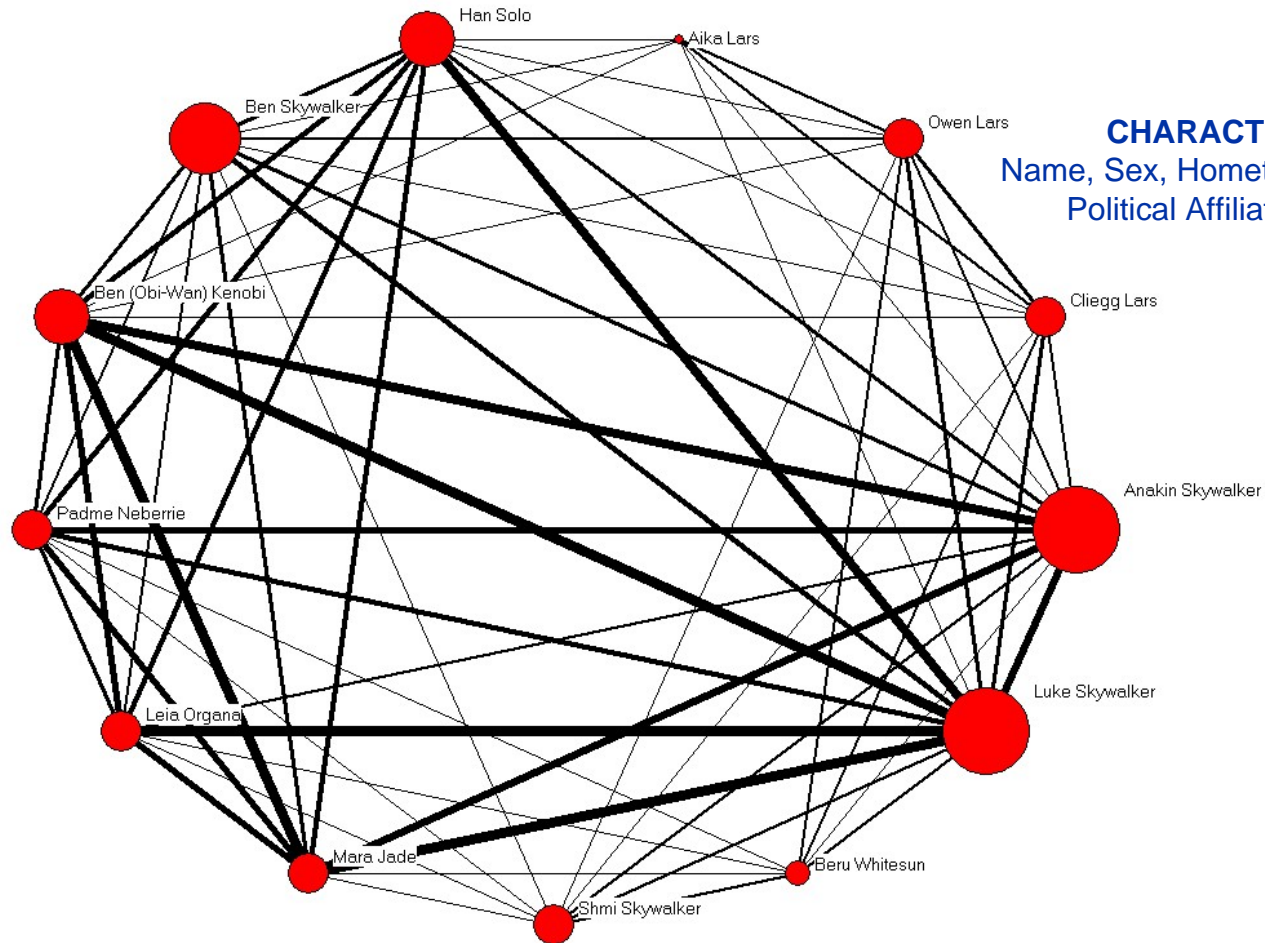


(Source: [http://en.wikipedia.org/wiki/Skywalker\\_family](http://en.wikipedia.org/wiki/Skywalker_family))



# Total Affinities

(Thicker lines indicate stronger affinities --- Highly connected group )



**CHARACTERISTICS**  
Name, Sex, Hometown, Occupation,  
Political Affiliation, Children

In this graph, the **larger nodes** indicate that the individual has **more affinities** with others.  
(i.e., Luke Skywalker , the largest, has affinities with everyone; Aika Lars, the smallest, has affinities with seven others)

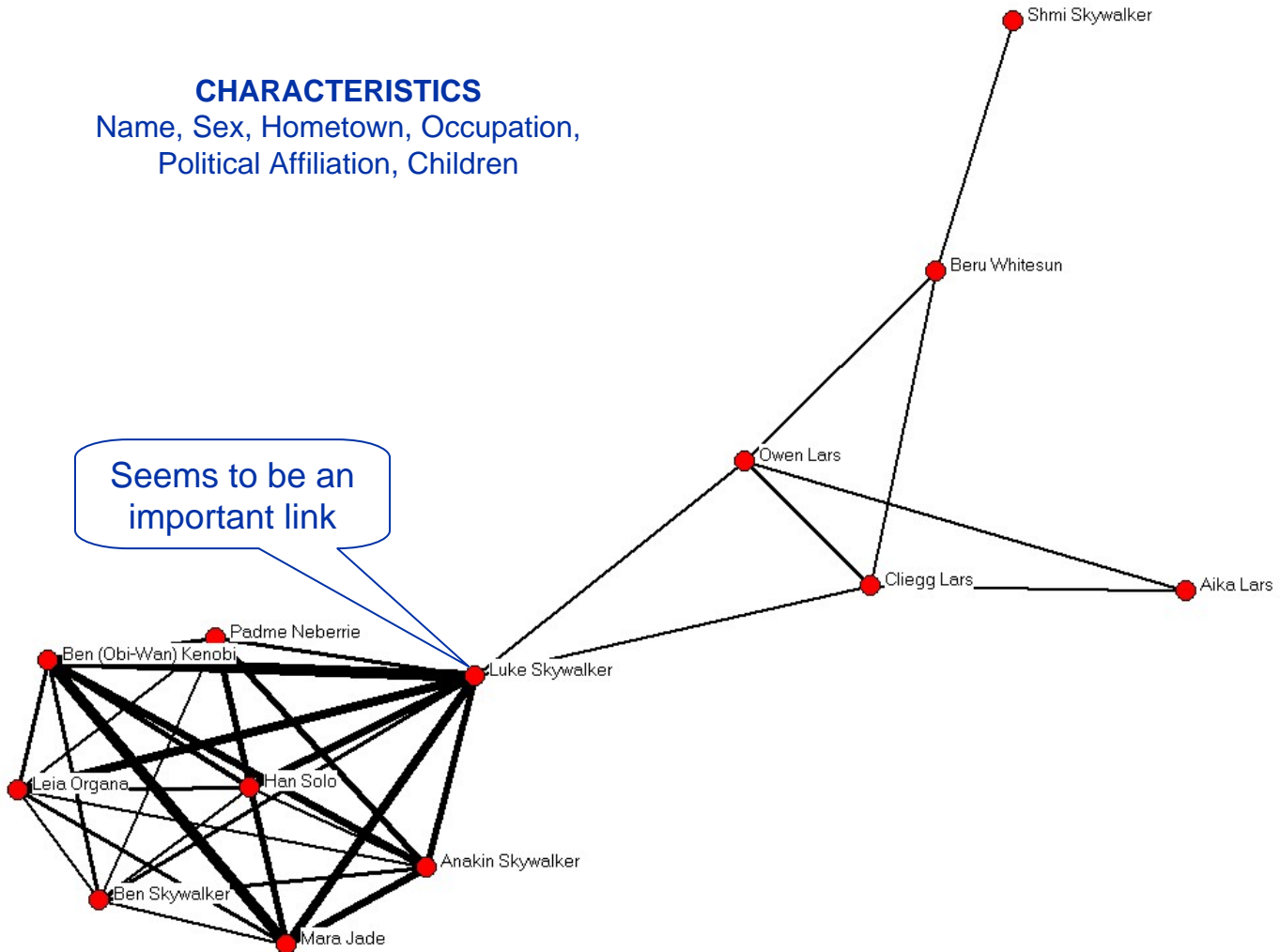


# Total Affinities

(More than two Affinities)

## CHARACTERISTICS

Name, Sex, Hometown, Occupation,  
Political Affiliation, Children

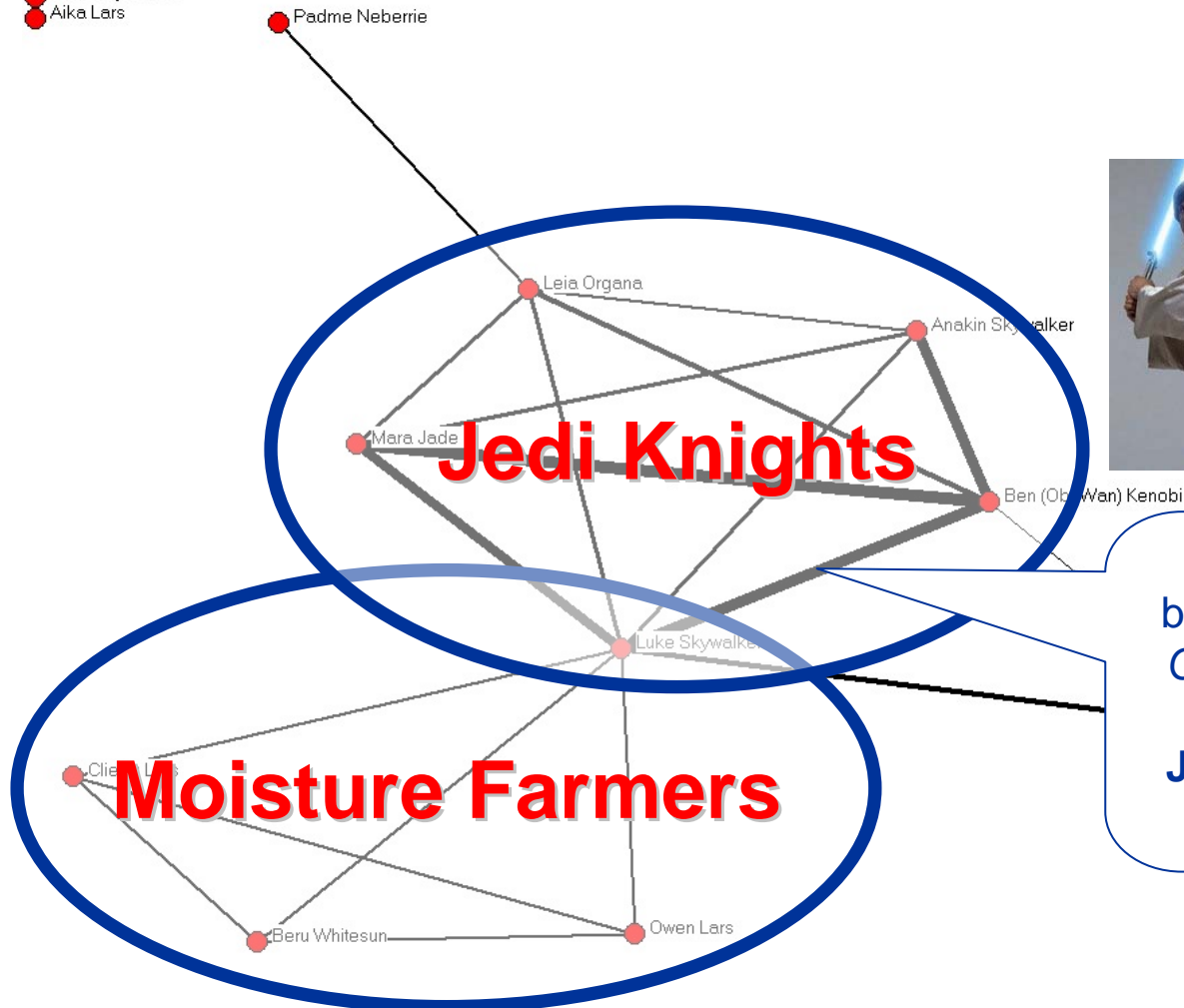


# Occupational Affinity Network

## Network Discoveries (Star Wars)



- Ben Skywalker
- Shmi Skywalker
- Aika Lars



Stronger Affinity between *Luke* and *Obi-Wan* because they were both **Jedi Knights** and **Jedi Masters**

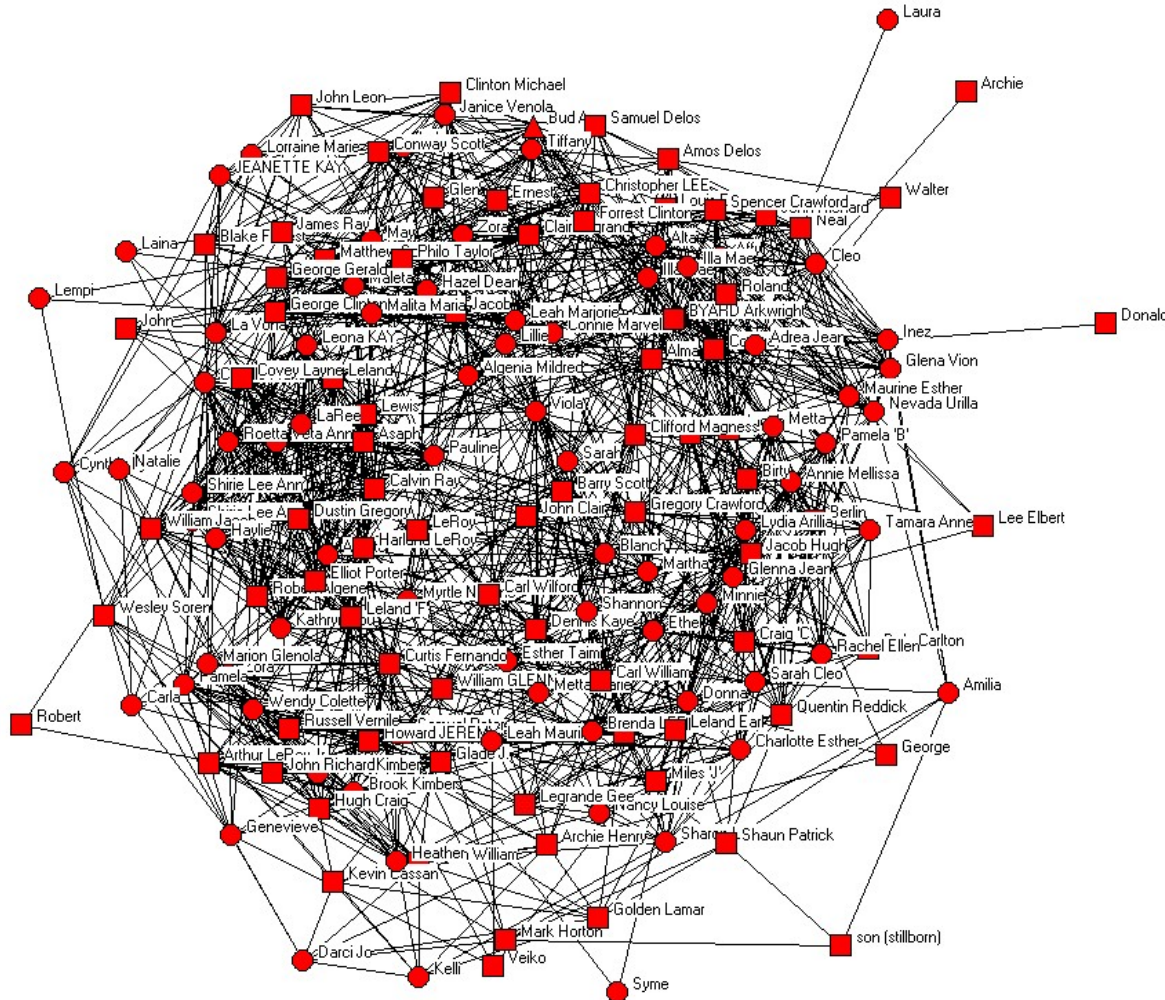


# Real Family Data

- Typical GEDCOM file
- That had **only basic information**:
  - Name
    - Given and surname
  - Sex
  - Birth Information
    - Date and place
  - Death Information
    - Date and place

# Birthday Network

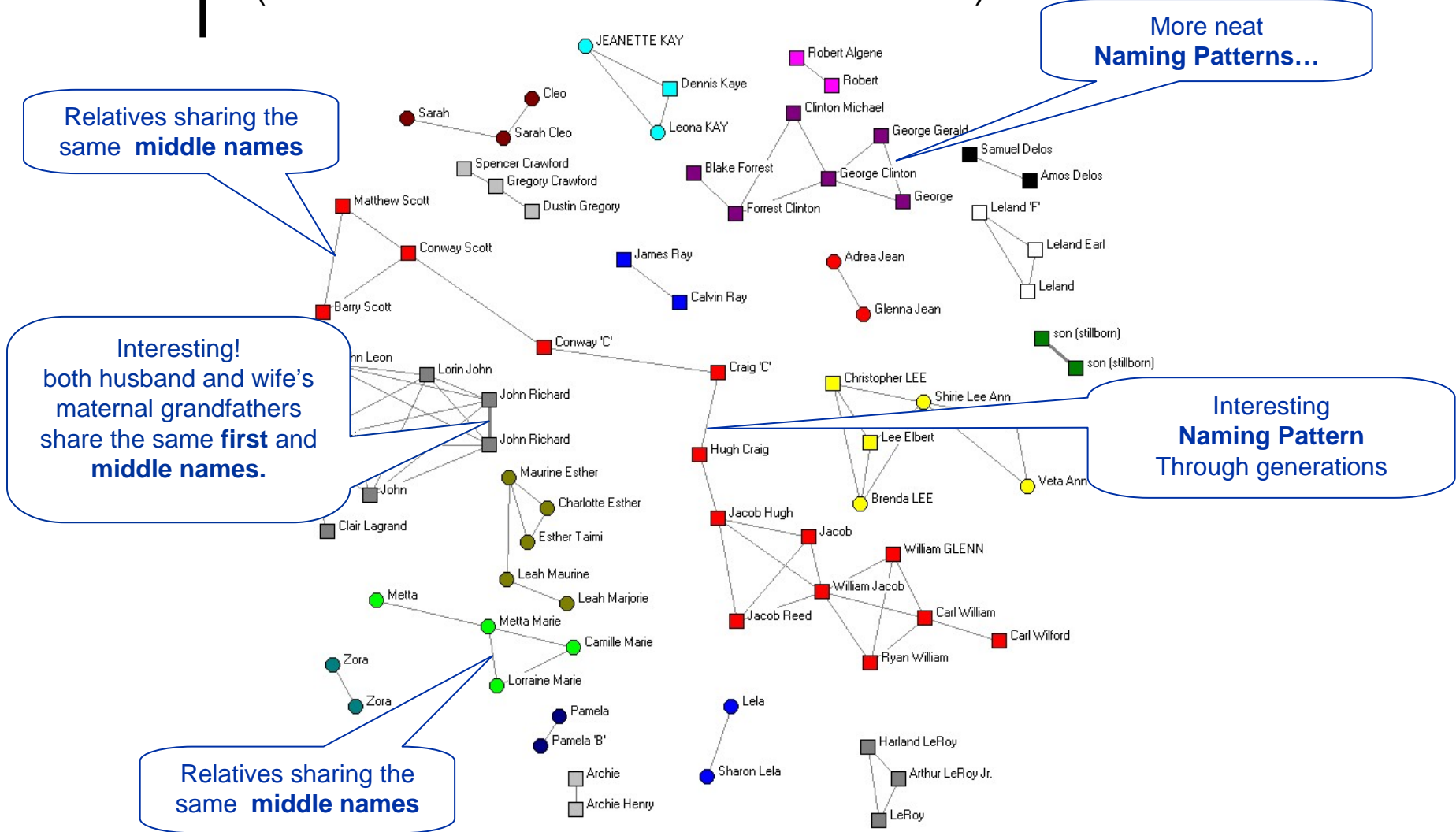
(One or more affinities) Somewhat difficult to interpret





# Given Name Network

(One or more affinities --- Isolates removed)





# Richer data is even better

- Even **basic family data** produced networks that **revealed interesting discoveries**
- Clearly, the **richer the data**, the **more interesting** the affinity networks
- What data supports interesting affinity networks?
  - What affinities are interesting to your family?
  - Is family member geography important?
  - Are family members' interests and hobbies important?
  - What social aspects of life are of interest?
  - What occupational data might be useful?



# Other Attributes

(That would make interesting affinity networks)

- Education
- Physical traits – hair, eyes, height, weight, etc.
- Occupation – employment status, age of retirement
- Special achievements
- Hobbies, talents, and sports
- Ethnic or racial background
- Religion or religious change
- Military service and where served
- Dates of family members leaving or returning home
- Places lived
- **Anything else that is interesting to the family!**



# Recording Attributes

(Interestingly, the GEDCOM standard already allows for this)

- **GEDCOM 5.5 Tags** (Currently ~130):

ABBR, ADDR, ADR1, ADR2, ADOP, AFN, AGE, AGNC, ALIA, ANCE, ANCI, ANUL, ASSO, AUTH, BAPL, BAPM, BARM, BASM, BIRT, BLES, BLOB, BURI, CALN, CAST, CAUS, CENS, CHAN, CHAR, CHIL, CHR, CHRA, CITY, CONC, CONF, CONL, CONT, COPR, CORP, CREM, CTRY, DATA, DATE, DEAT, DESC, DESI, DEST, DIV, DIVF, DSCR, EDUC, EMIG, ENDL, ENGA, EVEN, FAM, FAMC, FAMF, FAMS, FCOM, FILE, FORM, GEDC, GIVN, GRAD, HEAD, HUSB, IDNO, IMMI, INDI, LANG, LEGA, MARB, MARC, MARL, MARR, MARS, MEDI, NAME, NATI, NATU, NCHI, NICK, NMR, NOTE, NPFX, NSFX, OBJE, OCCU, ORDI, ORDN, PAGE, PEDI, PHON, PLAC, POST, PROB, PROP, PUBL, QUAY, REFN, RELA, RELI, REPO, RESI, RESN, RETI, RFN, RIN, ROLE, SEX, SLGC, SLGS, SOUR, SPFX, SSN, STAE, STAT, SUBM, SUBN, SURN, TEMP, TEXT, TIME, TITL, TRLR, TYPE, VERS, WIFE, and WILL.

(Source: The GEDCOM Standard Release 5.5, Appendix A)

- **Other tags**, or attributes, not currently defined could be stored as notes (which can be text mined).



# Conclusion

- This presentation described a **method for building networks that highlight affinities** from family history data
- The content of such affinity networks can be exploited to **strengthen living families** and to **direct family history research**.
- Knowing that we have some affinity with ancestors **encourages us to find out even more** about them, bringing them “closer” to us and thus effectively **“turning our hearts to them.”**



# Contact Us:

Brigham Young University

Data Mining Lab

1138 TMCB

Provo, UT 84602

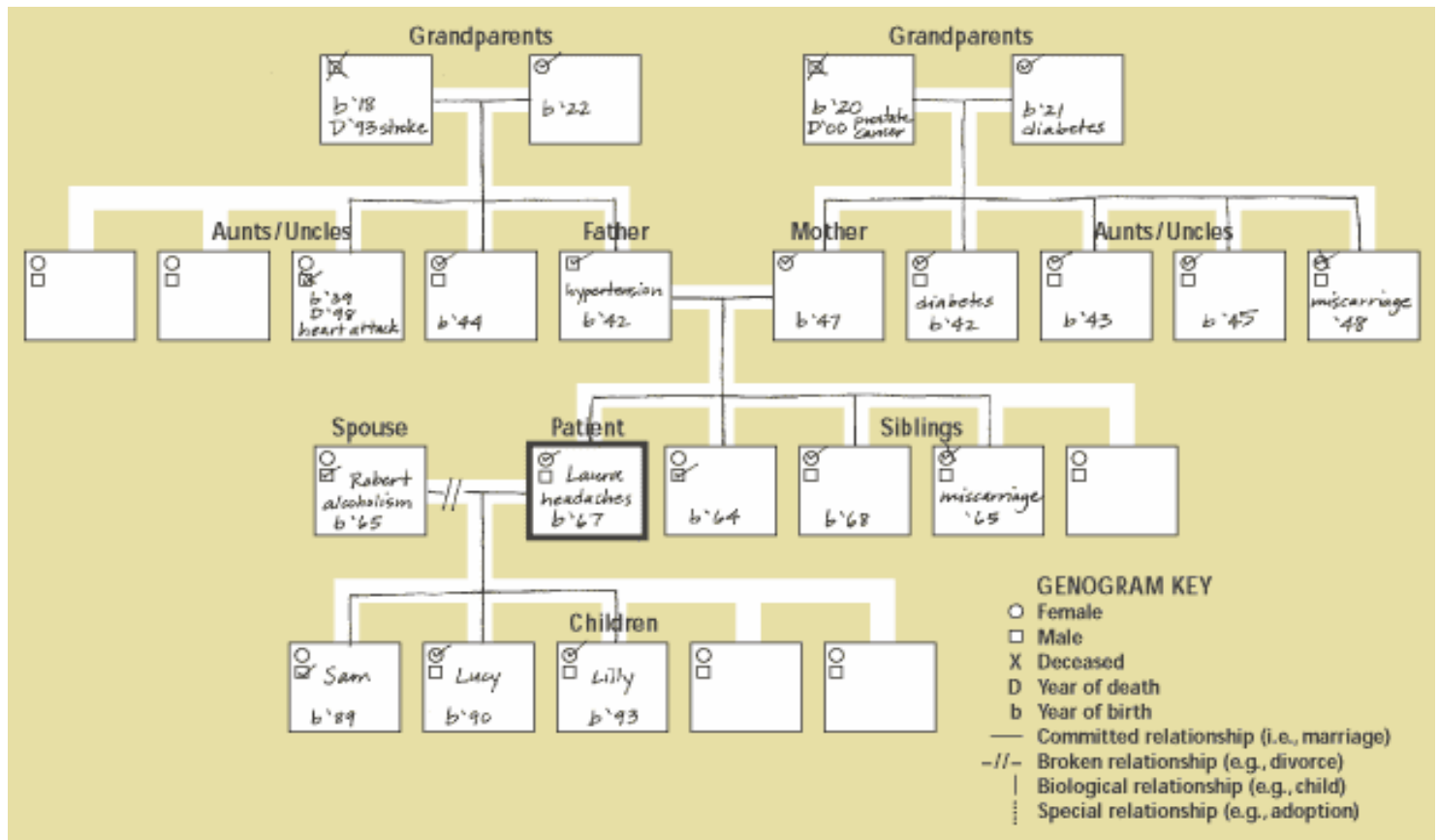
(801) 422-7817

[smitty@byu.edu](mailto:smitty@byu.edu)

We'd like to make this widely available

# Other Ideas:

(Record medical histories to build a medical affinity network)



Source: <http://www.aafp.org/fpm/20010300/49focu.html>