Characterizing HPV Vaccine Discourse On Reddit

AMIA SMM4H 2017

Yuki Lama, MPH[1]
Dian Hu, Doctoral Student[2]
Sandra C. Quinn, PhD[1]
Amelia Jamison, MA, MPH[1]
David A. Broniatowski, PhD[2]

[1] University of Maryland
[2] George Washington University

Background

◎ Human papillomavirus (HPV) is a common sexually transmitted infection that can lead to cancer

◎ Approximately 23,000 women and 16,500 men affected by HPV related cancers (CDC)

◎ HPV vaccination is recommended for preteens 11-12 years old as preventive measure

Increase the vaccination coverage level of 3 doses of human papillomavirus (HPV) vaccine to 80% for females and males aged 13 to 15 years*

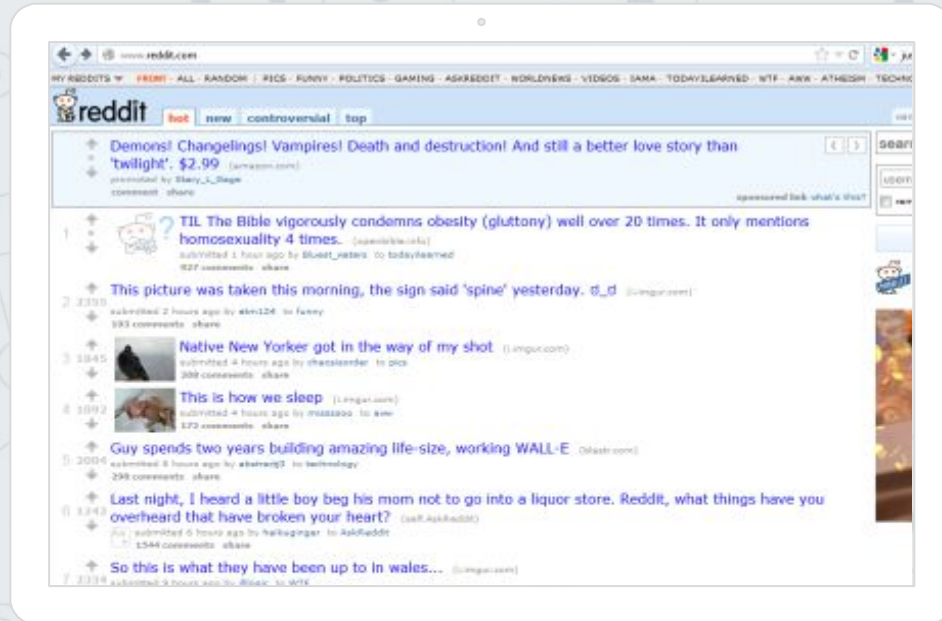**Healthy People 2020**

Public Health Communication Efforts

How are people accessing and using health information?

Social media as a tool for surveillance and health promotion.

How are users discussing HPV vaccine on Reddit?

◎ Cancer risk vs. sexual behavior concerns
◎ Temporal trends

DOMAIN:	DESCRIPTION:	EXAMPLE:
Cancer Risk	Messages are related to HPV and relationship to cancer and HPV vaccine. This includes concerns related to cancer, awareness of cancer risk. Comments can be any related type of cancer: cervical, throat, penile, anal, etc.	*it's not just genital warts, the hpv vaccine can prevent cervical cancer. amazing stuff.* *the vaccines don't prevent all causes of cervical cancer, they protect against some strains of hpv which can cause cervical cancer. also, they only started giving that vaccine to women over 25 recently.*
General Sexual Behavior	Messages are related to HPV and sexual activity including discussions of age of sexual debut, risky sexual behavior, multiple sex partners, associations with promiscuity, etc. Messages need an explicit connection to HPV. Can include screening behavior if it is related to sexual activity.	*seeing as he has already claimed that the hpv vaccine would make his daughters promiscuous and tried to stop it being made compulsory for girls, it would not surprise me at all.* *thanks for clearing that up. i'm a guy, so it's a little tough to wrap my head around. i might get the vaccine just as a preventative measure. i've only had one sexual partner in my history and we were both virgins, so i'm certain i'm clear. not that hpv is a death sentence or anything, but i'd rather not have to deal with it, and so i don't unknowingly spread it to others in the case that i did get it.*

Qualitative codebook and sample messages

Methods

1. Data Source:
    ○   All public Reddit comments from Jan 2006 to Dec 2015. 22750 potential HPV-vaccine related messages were identified that include the string "hpv" and "vaccin".

2. Annotation and Keyword selection:
    ○   In each iteration, two annotators manually evaluated approximately 100 of the messages using the qualitative codebook. Two iterations are performed to discover keywords with high importance.
    ○   In the third iteration, 400 messages are annotated to evaluate the keyword performance and used to build a classifier.

Methods

3. Categories:
   ○ (1)cancer risks or (2) sexual behavior

4. Building classifier:
   ○ 450 out of 600 messages are used to train the classifier.
   ○ 150 out of 600 messages are used for testing.
   ○ Along with bag of word vector as features, whether "keyword occurred" is used as one binary feature in the classifier.
   ○ Cross validation are performed using K=2 to 10.

Methods Performance

1. Performance of Keyword only method :

Out of the 200 messages identified as related to cancer only using keywords methods, **Annotator 1 identified that 187 messages** are actually related to cancer. **Annotator 2 identified that 178 messages to be related. (accuracy: 0.935/0.89)**

Out of the 200 messages identified as related to sexual behavior only using keywords methods, **Annotator 1 identified that 122 messages** are actually related to sexual behavior. **Annotator 2 identified that 160 messages to be true. (accuracy: 0.61/0.8)**

Methods Performance

2. Performance of two contemporary classifiers:

Sexual Activity Classifier, (K=5):
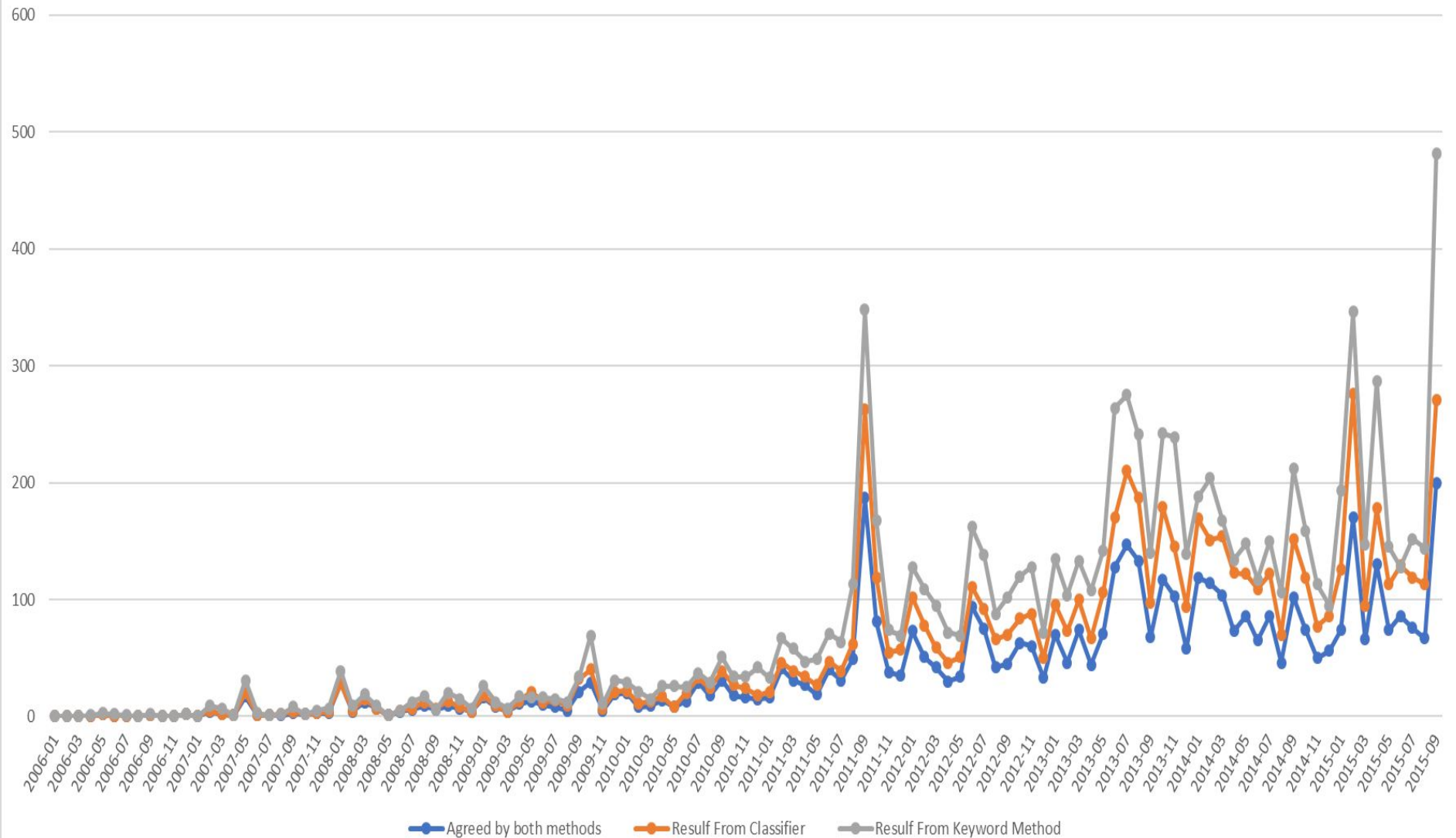precision recall   f1-score support
0.72   0.71   0.71   150

Cancer Classifier, (K=5):
precision recall   f1-score support
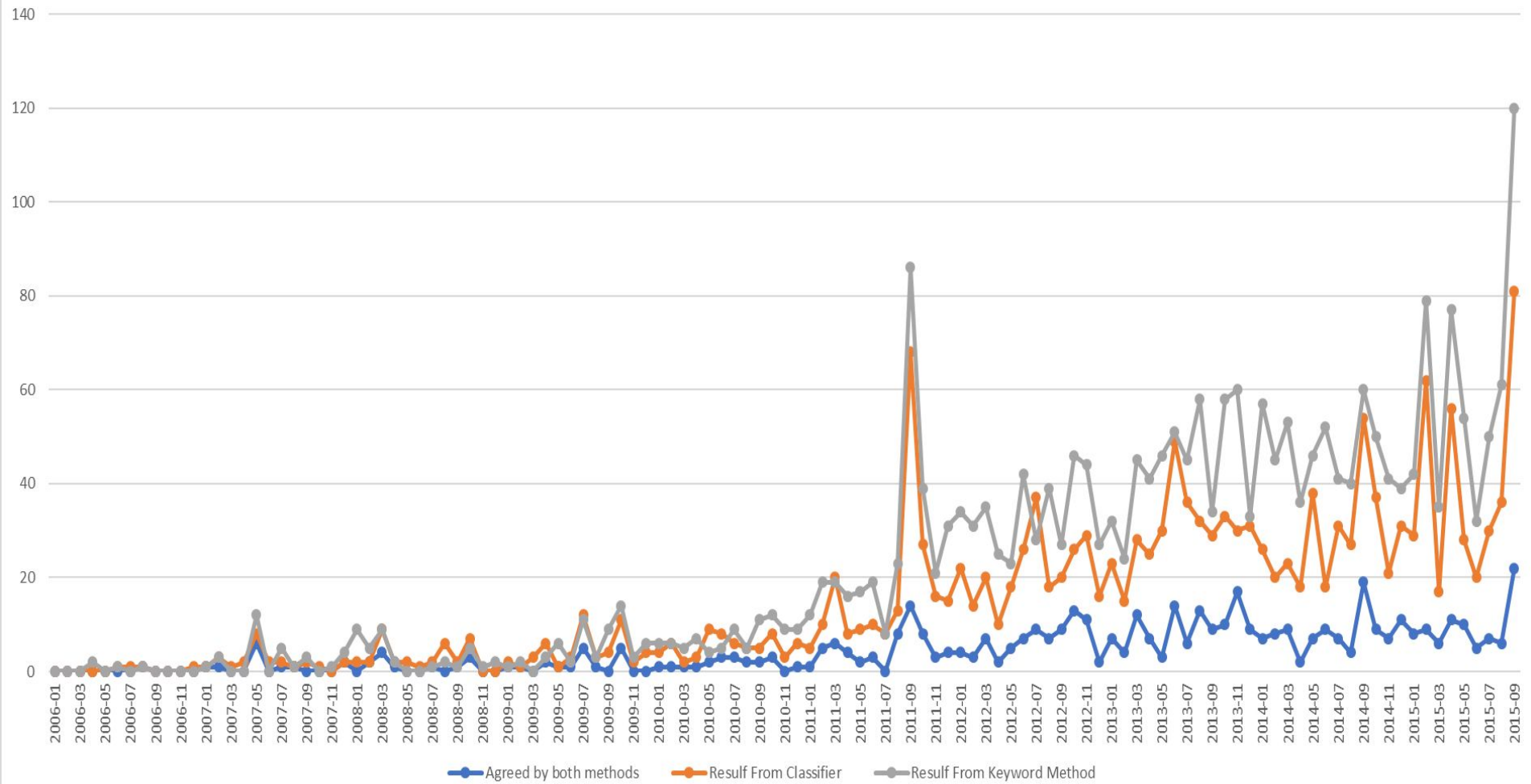0.73   0.57   0.61   150

Preliminary Results

Monthly Frequency of HPV vaccine related messages, discussing risk of cancer

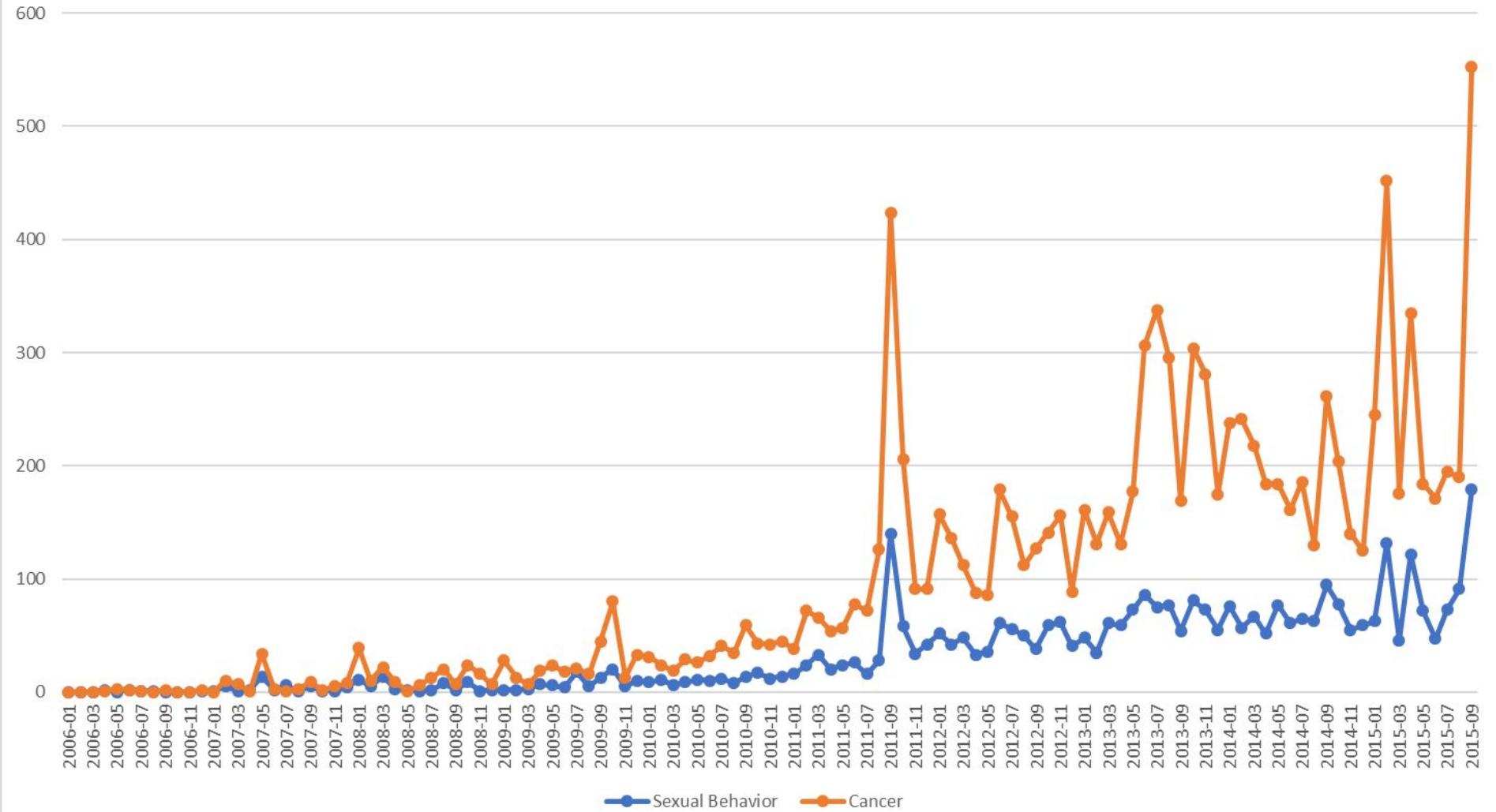**– Agreed by both methods – Results from classifier – Result from keyword method** **12**

Preliminary Results

Monthly Frequency of HPV vaccine related messages, discussing sexual behavior

– **Agreed by both methods**   – **Results from classifier** – **Result from keyword method**   **13**

Preliminary Results

Sexual Behavior v.s. Cancer, Messages labeled as related by at least one method

— Sexual behavior   — Cancer

Discussion

◎ Increased discussion of HPV vaccine

◎ Cancer predominantly being discussed

◎ Seminal timepoints
- October 2011 (vaccine introduced for boys)

◎ Limitations

◎ Future directions

Thank you!

You can contact us at:

lamay@umd.edu

hudian@gwmail.gwu.edu