**1. Replicate 2012-xiang**

**Features:**

* Factual Features (CrunchBase, Dec-2013)
  + Basic Features e.g. number of employees, office location, company age
  + Financial Features e.g. number of funding rounds, investment per funding round
  + Managerial Features e.g. number of acquired companies by founders
* Topic Features (TechCrunch articles, up to Dec-2013)

**Outcome:** Acquired? – CrunchBase (Dec-2013)

**Processing:**

* Topic model - Latent Dirichlet Allocation (LDA)
* Classification
  + Bayesian Network (BN)
  + Support Vector Machines (SVM)
  + Logistic Regression (LR)

**2. Extend binary classification techniques**

* Random Forest (RF) – as in 2016-yuan
* CART Decision Tree (CART) – as in 2016-beckwith
* Restricted Bolzmann Machine (RBM) – as in 2016-beckwith

**3. Extend factual features**

* Twitter presence (CrunchBase, Dec-2013) – as in 2016-beckwith
* Founder education (CrunchBase, Dec-2013) – as in 2016-beckwith
* Founder employment (CrunchBase, Dec-2013) – as in 2016-beckwith
* Founding team (CrunchBase, Dec-2013) – as in 2013-spiegel

**4. Extend problem to multiclass classification**

* Potential outcomes: Bankrupt**,** Acquired**,** IPO**,** Continue
* Potential multiclass classification techniques: One vs. all (OVA), All vs. all (AVA)

**5. Extend problem to longitudinal study**

* Select continuing businesses at Dec-2013
* Use features at Dec-2013 to predict outcome at Jul-2016
* Use features at Jul-2016 to predict outcome at Jul-2016

**6. Extend topic model processing techniques**

* Domain-Constricted LDA model – as in 2016-yuan

**7. Extend topic and network features**

* Patent similarity (Google Patents, up to Dec-2013) – as in 2015-huang
* Social network link prediction (CrunchBase, Dec-2013) – 2014-shi, 2013-liang\_yuan
* H1. Xiang's results can be replicated
* Xiang: [(2012, 2012), (2012, 2012)]
* Present: [(2012, 2012), (2012, 2012)]
* Develop an improved model
* H2. Xiang's classifiers can be beaten
* H3. Xiang's results can become more robust
* H2. Newer datasets will be more predictive
* [(2012, 2012), (2012, 2012)]
* [(2013, 2013), (2013, 2013)]
* [(2016, 2016), (2016, 2016)]
* H3. Classifiers will be stable over time
* Chi Square (Training, Test)
* [(2012, 2012), (2012, 2012)]
* [(2012, 2012), (2013, 2013)]
* [(2012, 2012), (2016, 2016)]
* [(2013, 2013), (2012, 2012)]
* [(2013, 2013), (2013, 2013)]
* [(2013, 2013), (2016, 2016)]
* [(2016, 2016), (2012, 2012)]
* [(2016, 2016), (2013, 2013)]
* [(2016, 2016), (2016, 2016)]
* H4. Newer classifiers will be more stable over time
* ANOVA (Training)
* [(2012, 2012), (2012, 2012)]
* [(2012, 2012), (2013, 2013)]
* [(2012, 2012), (2016, 2016)]
* [(2013, 2013), (2012, 2012)]
* [(2013, 2013), (2013, 2013)]
* [(2013, 2013), (2016, 2016)]
* [(2016, 2016), (2012, 2012)]
* [(2016, 2016), (2013, 2013)]
* [(2016, 2016), (2016, 2016)]
* 4. Predictive power through time
* [(2012, 2013), (2012, 2013)]
* [(2012, 2016), (2012, 2016)]
* [(2013, 2016), (2013, 2016)]