

R을 이용한 게임 데이터 분석

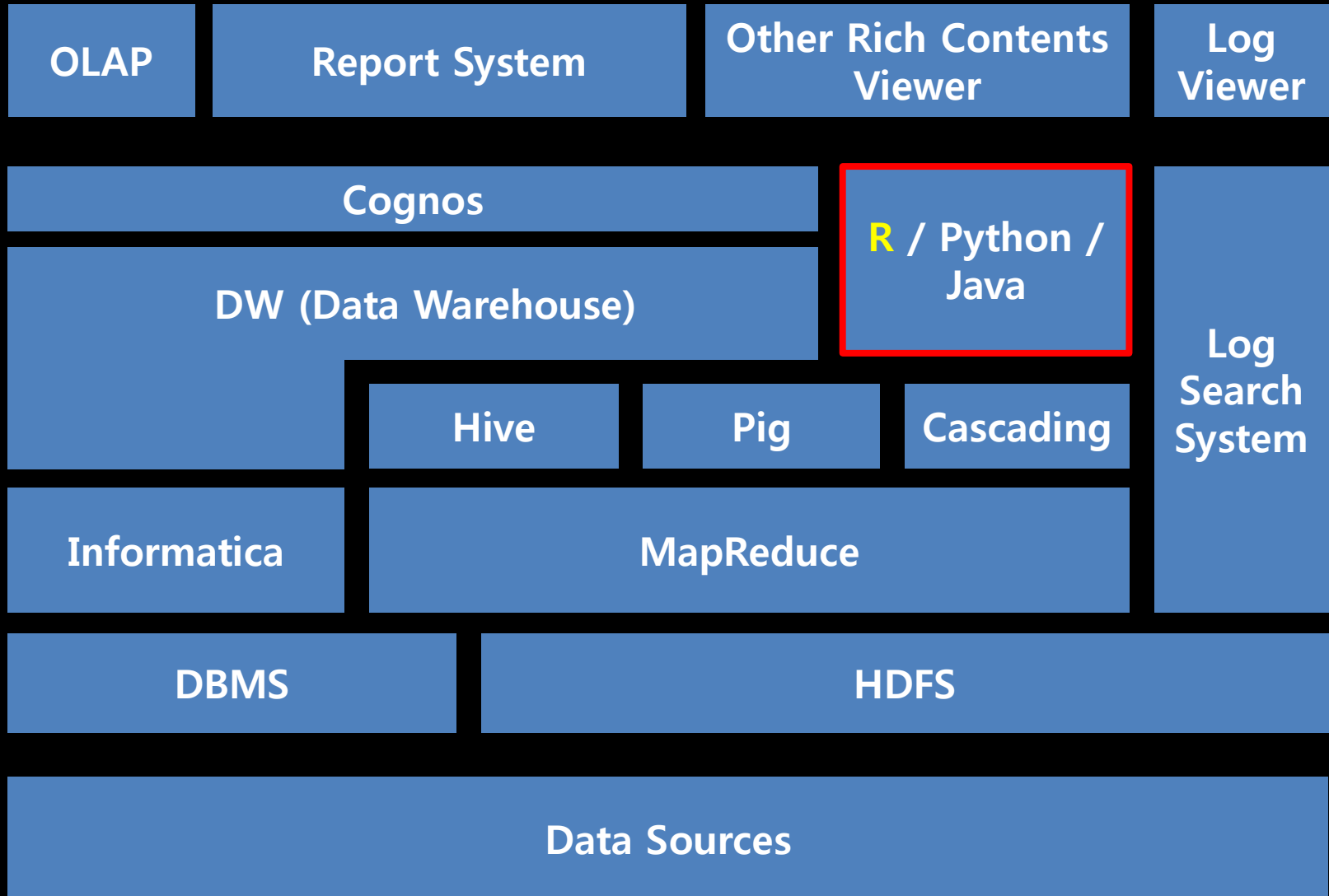
NCSOFT 데이터플랫폼실 GFIS팀

이은조

NCSoft

- 주요 제품
 - 리니지, 리니지2, 아이온, Blade & Soul, Guild War2
- MMORPG(Massively Multiplayer Online Role Playing Game)
 - 여러 유저간 다양한 **사회 활동**
 - 지속적인 **업데이트** 및 **고객 관리** 필요
 - 여러 가지 유료 상품 및 이벤트를 통한 **2차 수익 창출**

정보 처리 스택



GFIS

- Game **Fraud Investigation** Service
- Fraud Detection
 - 카드 도용
- Game Fraud Detection
 - 어뷰징
 - 계정 도용
 - 작업장



대박 터지는 매니아 쇼핑 찬스!!

시작페이지 > 즐겨찾기 >

이벤트 | 마이존 | 고객센터 | 서비스안내



NO.1 아이템거래 게임매니아 상품권몰 게임몰 매니아옥션 매니아존 커뮤니티 쇼핑

사용한 해피머니 상품권 있으세요?

최대 200만원

사용한 해피머니 상품권 핀번호 등록하고, 100% 즉시 당첨의 행운을 누리세요!

200만원 받으러 가기

레알팝 출시기념!! | 도서문상 마일리지 | 디

NO.1 아이템매니아

다이블로3, 리니지, 리니지2, 아이온, 던파, 메이플스토리 등 안전한 게임아이템 거래

19 최저가아이템 구매하기

게임몰

모든 게임 결제의 할인!

게임문 홈 >

NOTICE · [오픈] 레알팝 출시기념 이벤트

MEMBER LOGIN

아이디

와우 판다리아의 안개 업데이트!

와우 거래 1위 itemBay! 선착순 4,000명에게 계정비를 지원합니다!

Live바로가기 구매직접금 2배! | 즐겨찾기 | 아이템베이 앱 | 스마트폰 보안 | 고객센터 | 이벤트/쿠폰 | 게임몰리스 | 로그인

itemBay 바로고 안전한 아이템거래 NO.1

NO.1 아이템거래 마이베이 상품권/아이템샵 모바일세상 게임포털 전체메뉴 >

관심물품 빠르게 검색하기

게임검색

서버검색

팝니다 Q

삽니다 Q

판매 등록 >

구매 등록 >

팔고 싶어요! 판매등록 하기

사고 싶어요! 팝니다 검색하기

거래내역

정보관리

마일리지

메시지

합인쿠폰 등록

아이디 비밀번호 로그인

아이디 저장

아이디/비밀번호 찾기

와우 4번째 대규모 업데이트 "판다리아의 안개" 역대 최대규모!! 와우 계정비 4,000명에게 지원합니다.

베스트 게임 | 모바일/소셜게임 | 핫이슈 | 제휴이벤트

게임 시세 앱(APP)

게임 시세 앱(APP)

할인/혜택 서비스

↓ 즉석쿠폰! 최대 1만원 할인 >

↓ 신규가입 시 9,000원 할인 쿠폰 제공! >

2 Live 바로가기 방문 시 구매 직접금 2배 지급! >

게임별 거래 시세

아이템거래 GO >

최저가시세 >

평균시세 >

게임거래순위 >

200% 보상 | 보안서비스

다이블로3	시열 [마족]	7,087	▲ 223
리니지			
블레이드앤소울			
> 아이온			
리니지2			
던전앤파이터			

베는 다르다!

사고금액 200% 전액 보상제

상품권/아이템 할인샵

게임포털 온게임트

공지사항

베이뉴스

각종 상품권을 5% 싸게 구매!

SALE

문화 상품권 | 도서문화 상품권 >

하반기 OBTS 실시

첫경험을 준비하라

이벤트 기간: 2012.10.09 (화) ~ 2012.10.31 (수)

- [당첨자발표] 오프라인쿠폰 가입이벤트...
- [당첨자발표] 레알팝 정식오픈 축하 이...
- 플래시 게임 베스트 랭킹 마일리지 지...
- [당첨자발표] 카드당. 무조건 내지 말고...
- [당첨자 발표] 1,000만원 배 흥놀이 하...

적립금 2배!

마일리지 충전

마일리지 출금

메시지 보기

구매용 유통

판매용 유통

최근본 게임

경색한 게임이 습니다.

TOP

고객감동센터 1644-3333

와우거래 1위 itemBay

팀 업무 현황

- 데이터 분석과 서비스 개발 병행
- 데이터 분석 프로세스
 - Raw Data -> 정제/가공 -> 분석/모델링
 - 정제/가공: Pig
 - **분석/모델링: R**
 - 전체 프로세스 제어: Python
- 서비스 개발: MapReduce, Legacy Java, Python

R 사용 현황

- 탐지 로직 개발을 위한 사전 분석 목적
- Subversion Repository 연동
- R Studio 사용

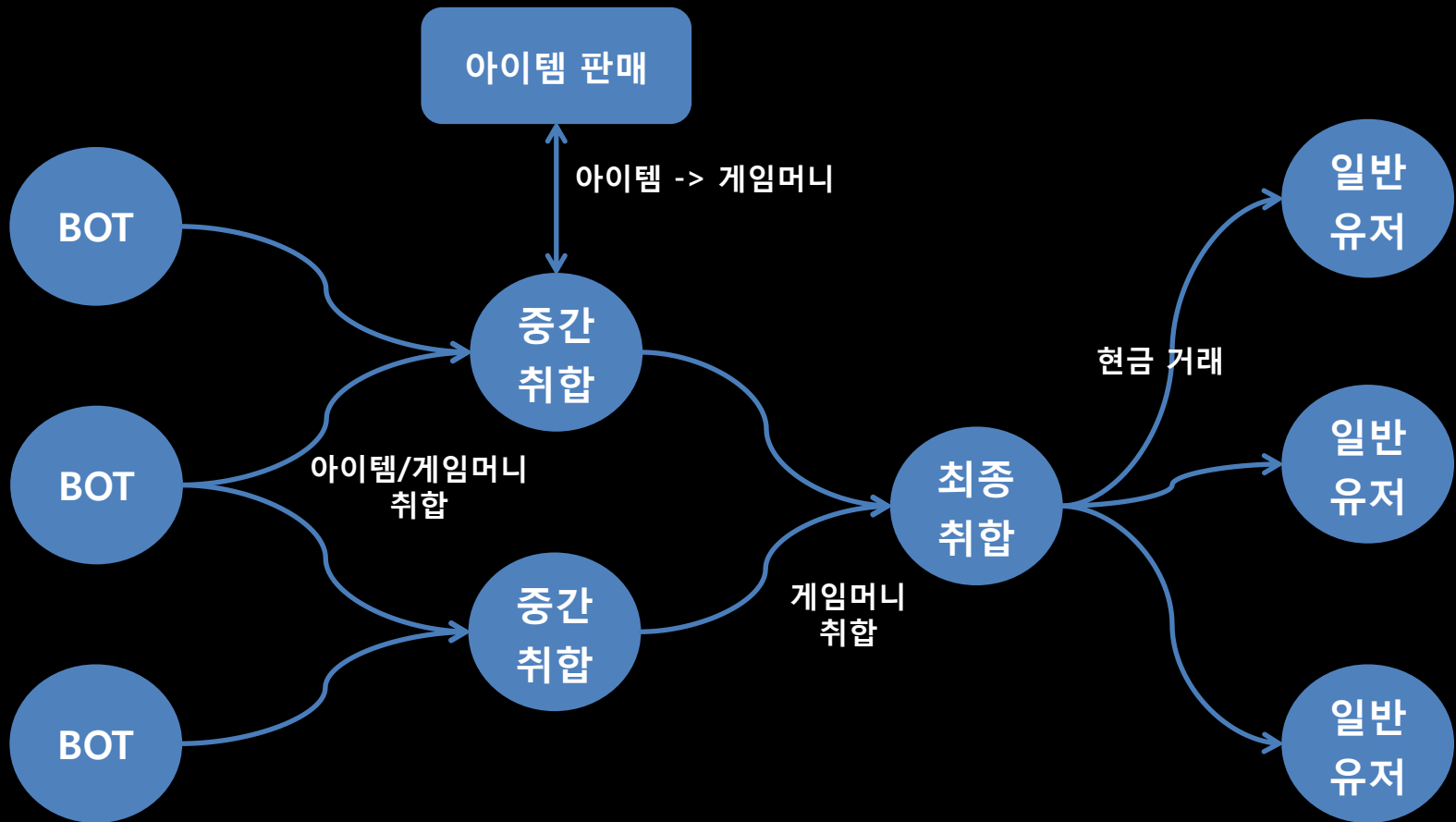
File/Folder	Size	Last Commit	Commit Message
trunk		7704 3 days	jslee509: #728 mod: pig 스크립트에서 날짜/시간 리스트 입력 파일 수정
bin		7704 3 days	jslee509: #728 mod: pig 스크립트에서 날짜/시간 리스트 입력 파일 수정
bin		7512 4 weeks	wonjun: #BYPASS SVN 소스 이전
gamelog2rdata.sh	2.3 KB	7512 4 weeks	wonjun: #BYPASS SVN 소스 이전
data		7557 2 weeks	wonjun: #729 mod: 상관행렬 데이터 처리된 RData 파일
data		7557 2 weeks	wonjun: #729 mod: 상관행렬 데이터 처리된 RData 파일
data		7557 2 weeks	wonjun: #729 mod: 상관행렬 데이터 처리된 RData 파일
data		7557 2 weeks	wonjun: #729 mod: 상관행렬 데이터 처리된 RData 파일
data	56.4 KB	7557 2 weeks	wonjun: #729 mod: 상관행렬 데이터 처리된 RData 파일
logs		7512 4 weeks	wonjun: #BYPASS SVN 소스 이전
pigs		7704 3 days	jslee509: #728 mod: pig 스크립트에서 날짜/시간 리스트 입력 파일 수정
batch_config.c3.py	11.1 KB	7640 6 days	wonjun: #728 add: 세그멘테이션 데이터 추출 Pig 스크립트 추가 및 batch_config 수정, constants.py 추가
batch_config.py	21.3 KB	7704 3 days	jslee509: #728 mod: pig 스크립트에서 날짜/시간 리스트 입력 파일 수정
batch_pig.py	2.1 KB	7640 6 days	wonjun: #728 add: 세그멘테이션 데이터 추출 Pig 스크립트 추가 및 batch_config 수정, constants.py 추가
constants.py	0.6 KB	7640 6 days	wonjun: #728 add: 세그멘테이션 데이터 추출 Pig 스크립트 추가 및 batch_config 수정, constants.py 추가
dump_gamelog_aggr.pig	1.0 KB	7704 3 days	jslee509: #728 mod: pig 스크립트에서 날짜/시간 리스트 입력 파일 수정
dump_gamelog_aggr_filter.pig	1.1 KB	7704 3 days	jslee509: #728 mod: pig 스크립트에서 날짜/시간 리스트 입력 파일 수정
dump_gamelog_aggr_range.pig	1.1 KB	7704 3 days	jslee509: #728 mod: pig 스크립트에서 날짜/시간 리스트 입력 파일 수정
dump_gamelog_cnt.pig	1.0 KB	7704 3 days	jslee509: #728 mod: pig 스크립트에서 날짜/시간 리스트 입력 파일 수정
dump_gamelog_cnt_filter.pig	1.1 KB	7704 3 days	jslee509: #728 mod: pig 스크립트에서 날짜/시간 리스트 입력 파일 수정
dump_gamelog_cnt_range.pig	1.1 KB	7704 3 days	jslee509: #728 mod: pig 스크립트에서 날짜/시간 리스트 입력 파일 수정
functions.py	0.9 KB	7512 4 weeks	wonjun: #BYPASS SVN 소스 이전
util.py	1.4 KB	7703 3 days	wonjun: #728 mod: 테스트 코드 주석처리
pyscripts		7696 3 days	wonjun: #728 mod: STL plot이 저장되는 디렉토리를 생성하도록 변경
rscripts		7692 4 days	jinalee: #727 r합수 수정(파일이름 받아오기)
analysis_regression_error.R	0.9 KB	7571 12 days	wonjun: #727 mod: Python 통합을 위해 rscript 경로를 pyscripts 하위로 변경
bns_corr.R	3.4 KB	7621 11 days	wonjun: #727 mod: ACF/STL 분석 함수를 common.R에서 분리
bns_rdata.R	7.3 KB	7602 12 days	jslee509: #727 mod: rscripts 경로 변경에 따른 bns_rdata.R의 내용 변경
bns_regress.R	9.0 KB	7685 5 days	jslee509: #729 bns_regress.R 수정
bns_stl.R	2.2 KB	7692 4 days	jinalee: #727 r합수 수정(파일이름 받아오기)

데이터 분석 사례

- 작업장 배후 계정 탐지
 - 네트워크 분석
- 이상 탐지
 - 자기 상관 분석
 - 회귀 분석

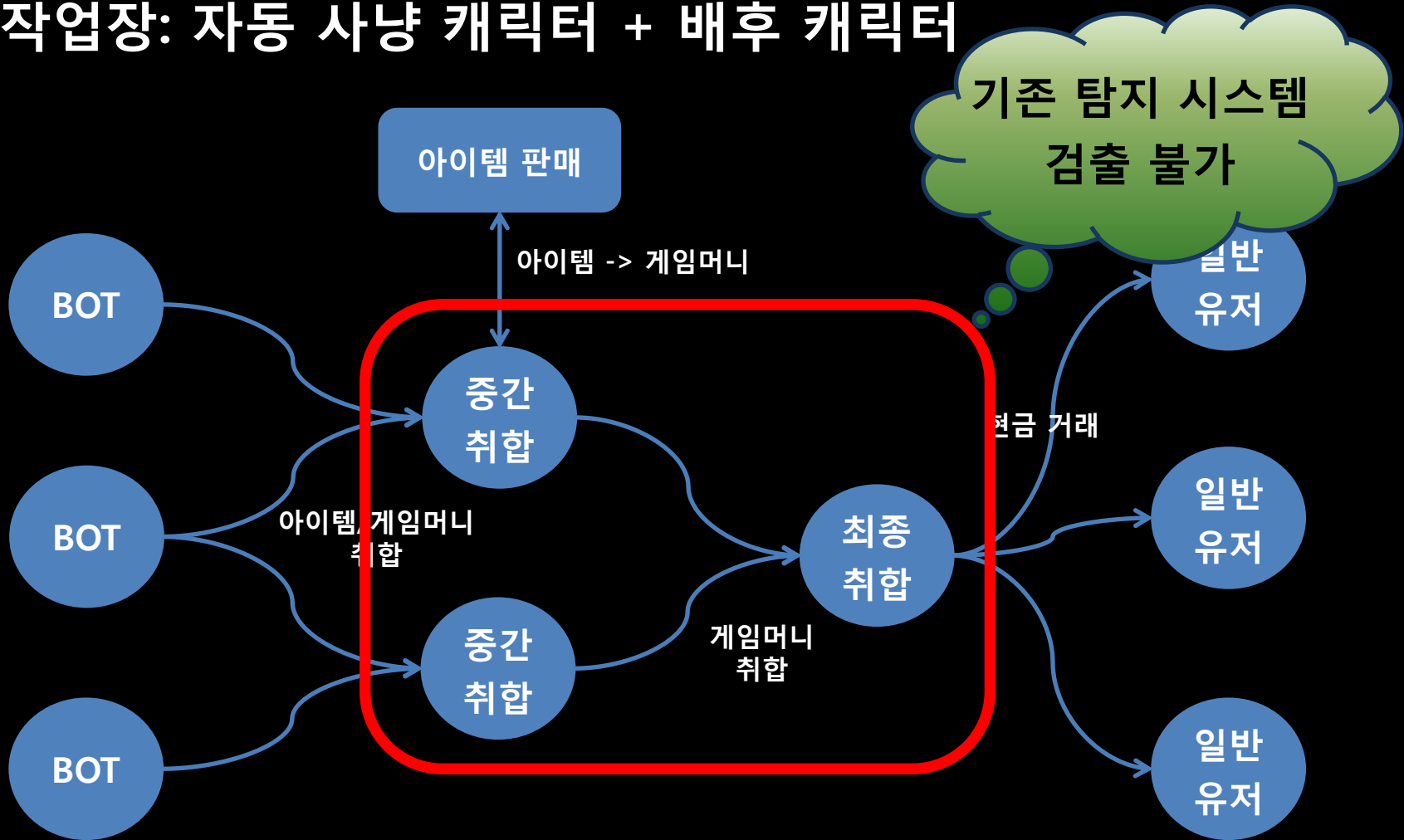
작업장 배후 계정 탐지

- 작업장: 자동 사냥 캐릭터 + 배후 캐릭터



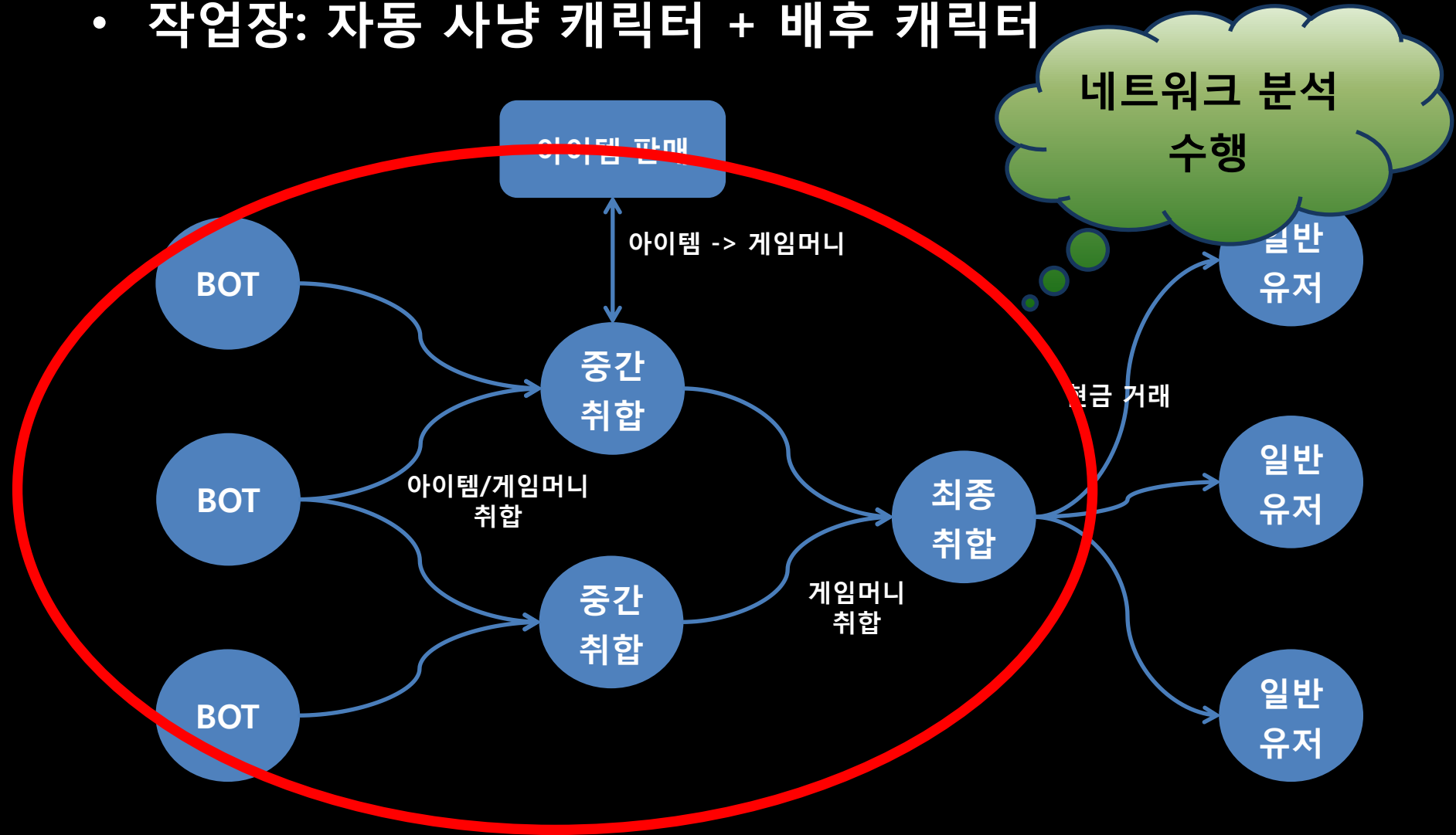
작업장 배후 계정 탐지

- 작업장: 자동 사냥 캐릭터 + 배후 캐릭터



작업장 배후 계정 탐지

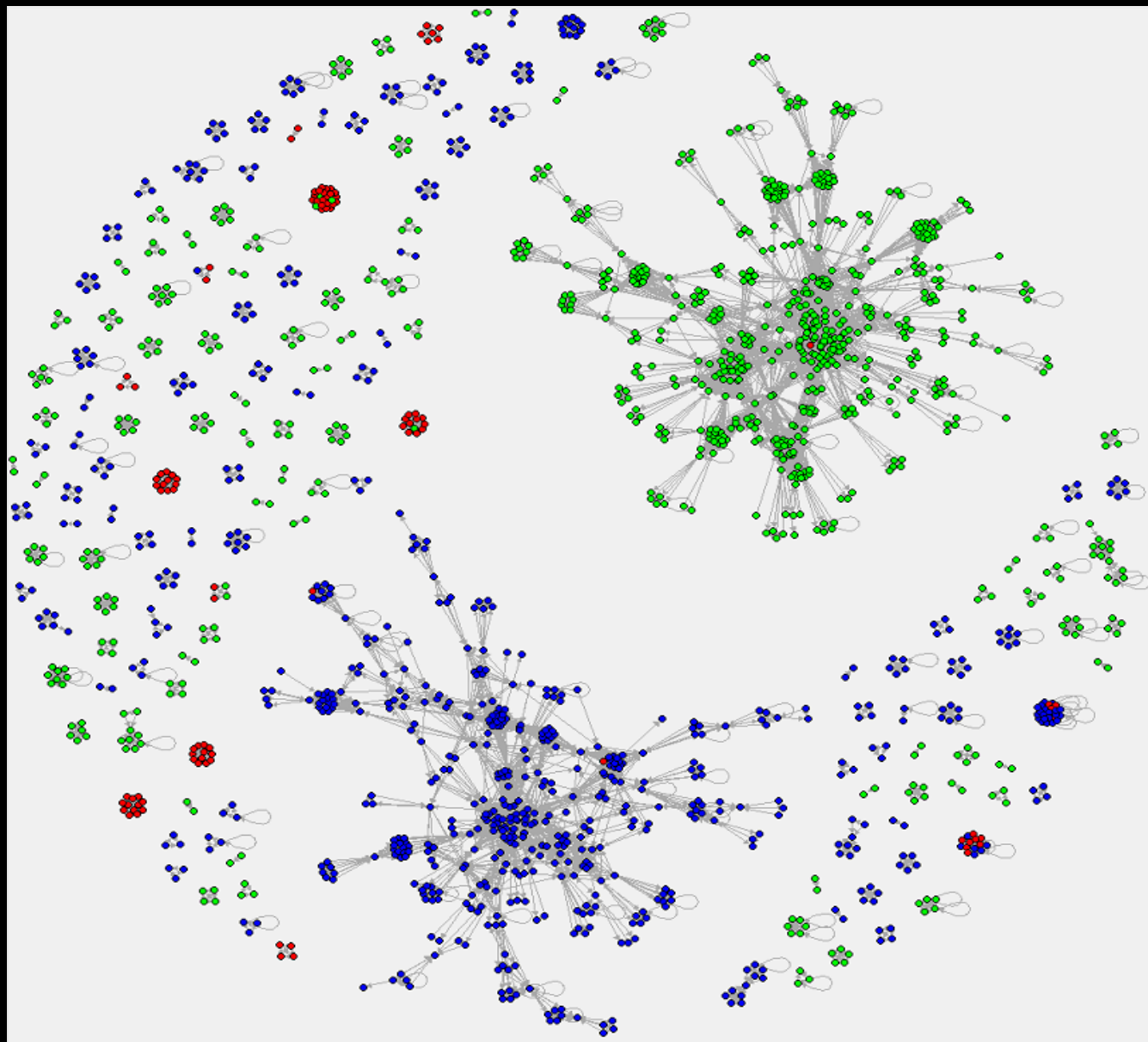
- 작업장: 자동 사냥 캐릭터 + 배후 캐릭터



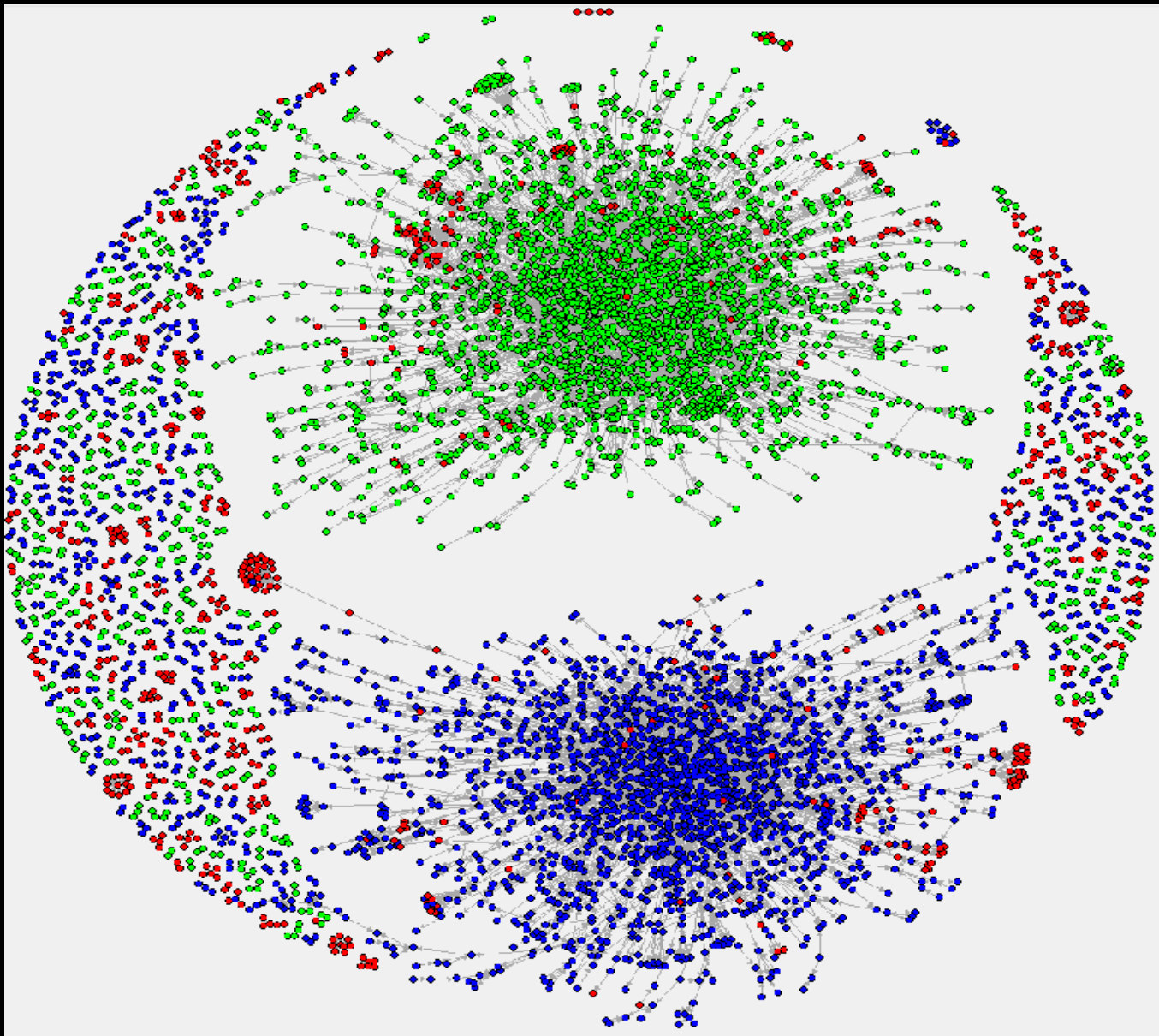
작업장 배후 계정 탐지

- 유저 데이터 분석 및 작업장 의심 캐릭터 추출
- 캐릭터 간 사회 활동 추출 -> **관계 네트워크** 구축
- **관계 네트워크 시각화**
 - > **비슷한 특성의 캐릭터끼리 모이는 경향** 확인

네트워크 시각화(파티)



네트워크 시각화(거래)



탐지 모델 구축

- Contagion Model

- 캐릭터 간 관계 및 불량성 **정량화**, **감염 로직** 적용

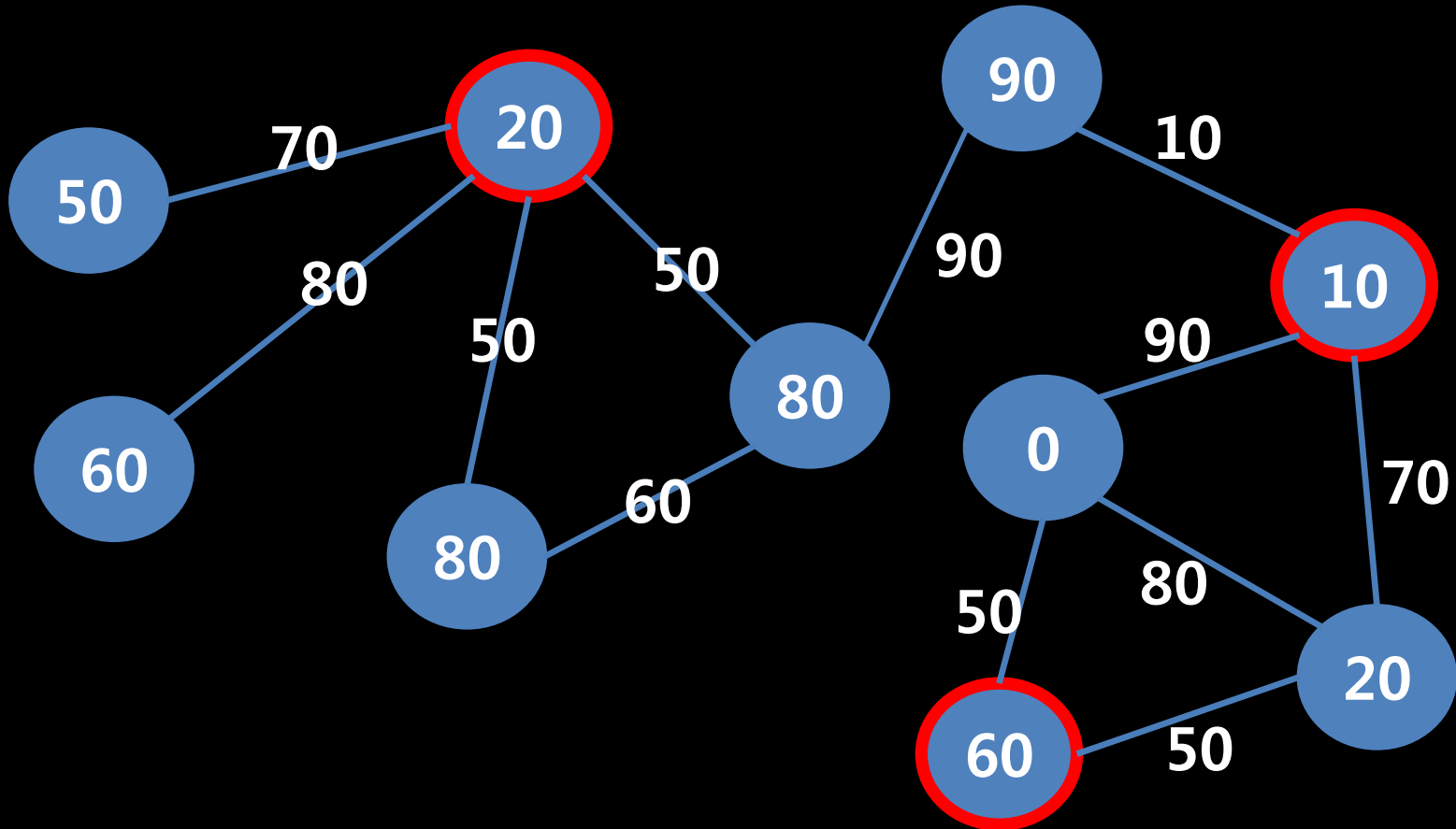
$$B(x) = B_0(x) + \sum_{f \in \text{Friend}(x)} \left(B_0(f) \frac{F(f, x)}{\sum F(f, f')} \right)$$

- Graph Clustering

- 긴밀한 네트워크 형성 **집단 분류**
- 작업장 여부 판별 기준 적용

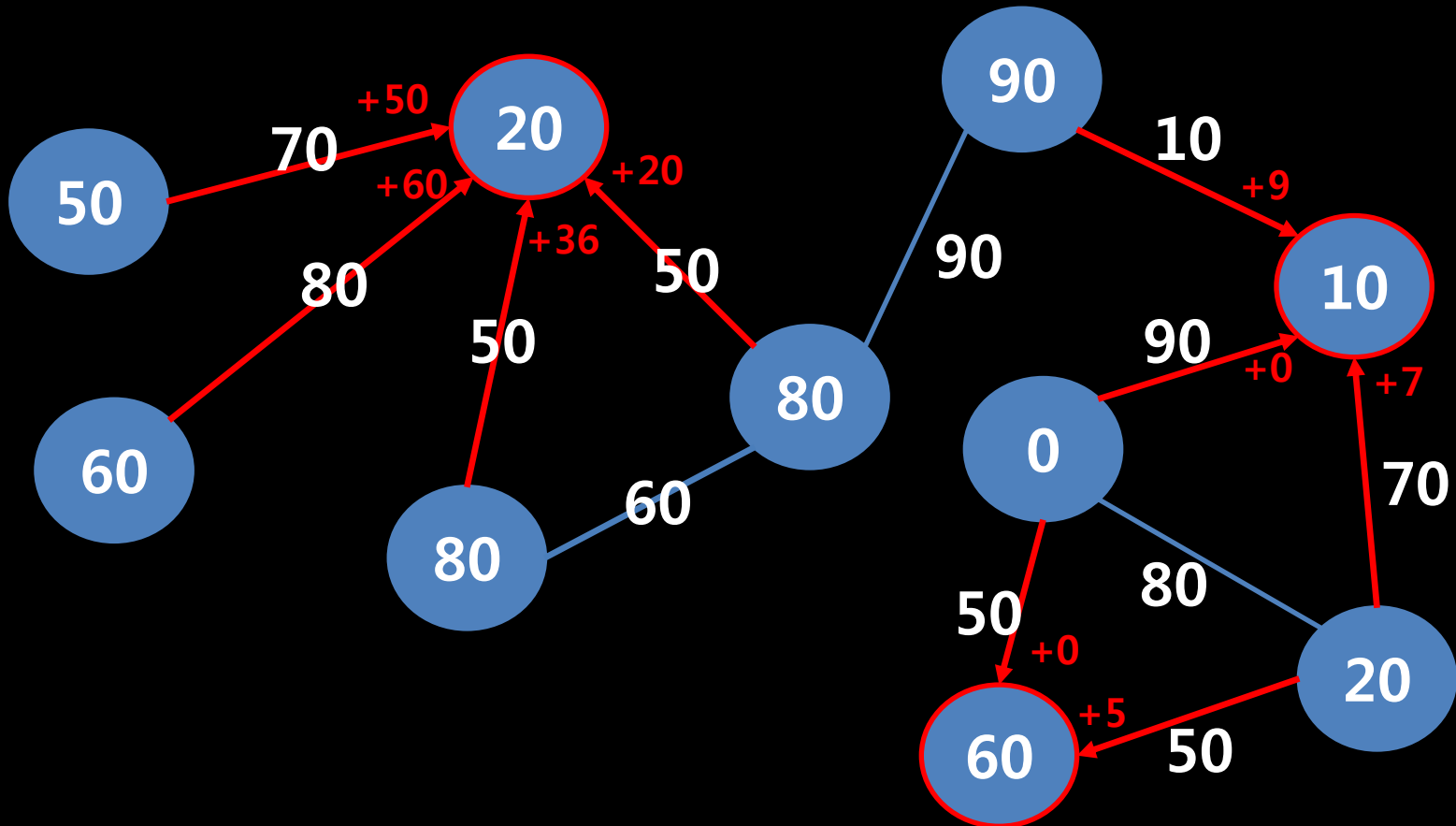
Contagion Model

$$B(x) = B_0(x) + \sum_{f \in \text{Friend}(x)} \left(B_0(f) \frac{F(f, x)}{\sum F(f, f')} \right)$$



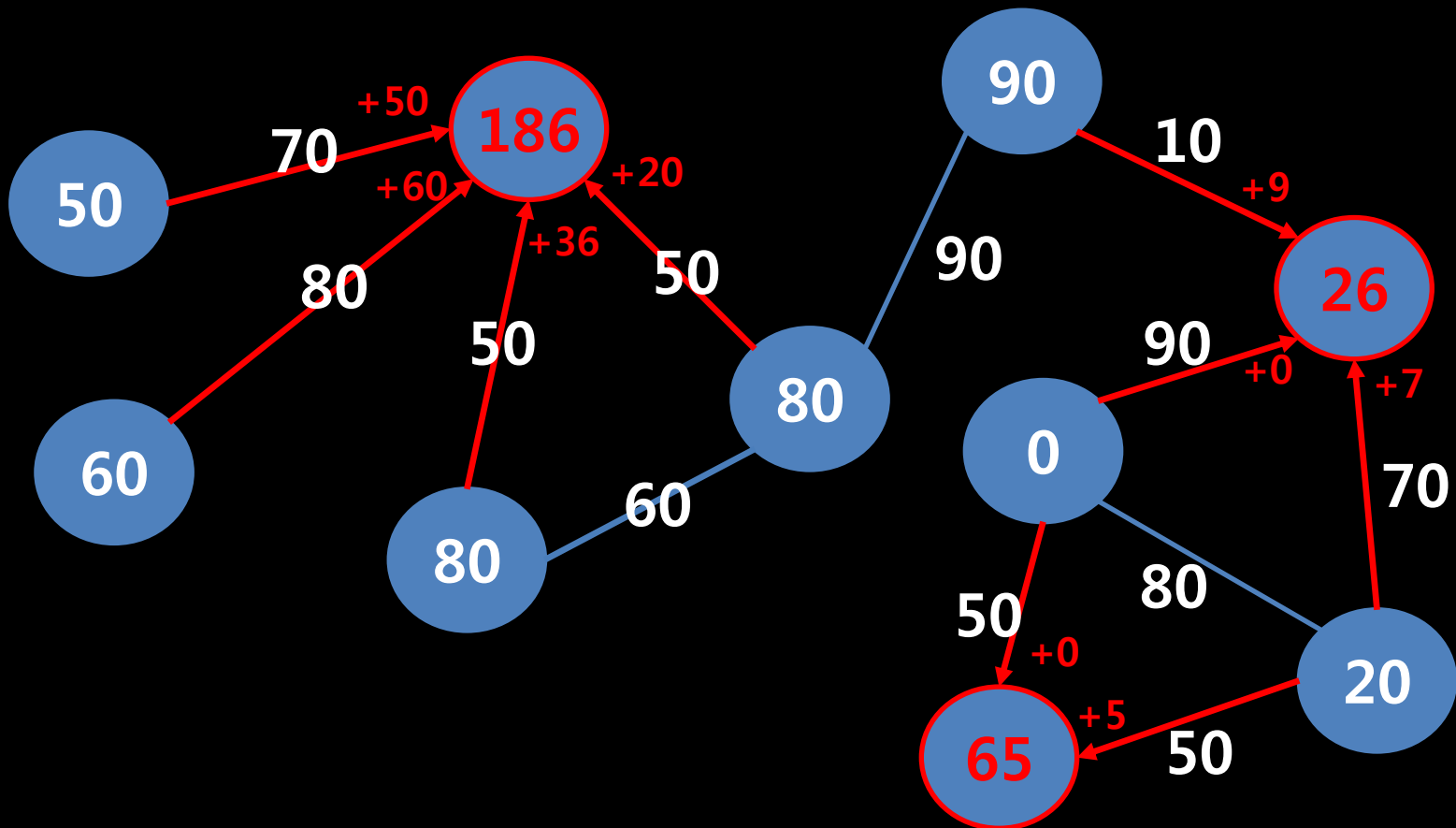
Contagion Model

$$B(x) = B_0(x) + \sum_{f \in \text{Friend}(x)} \left(B_0(f) \frac{F(f, x)}{\sum F(f, f')} \right)$$

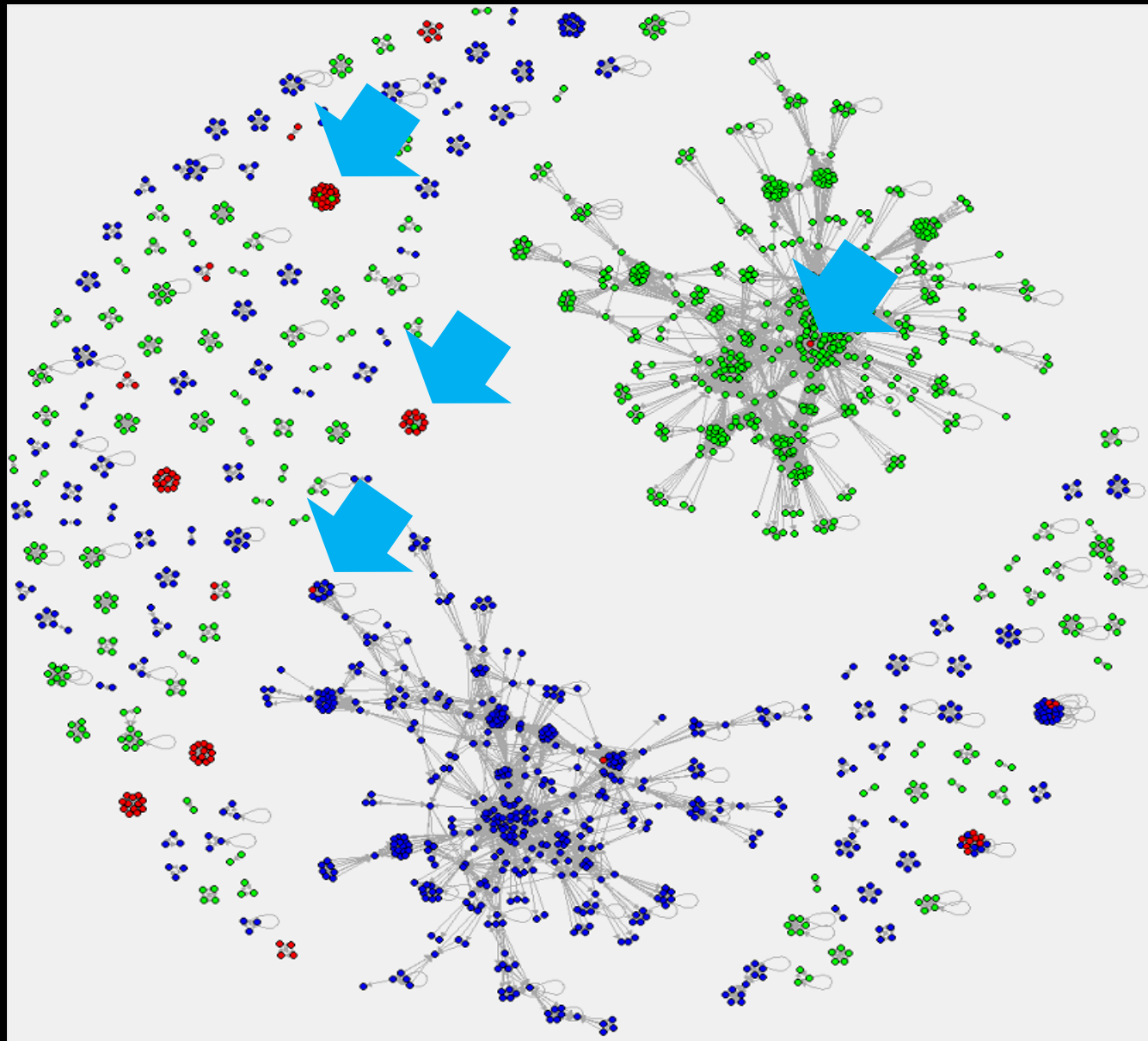


Contagion Model

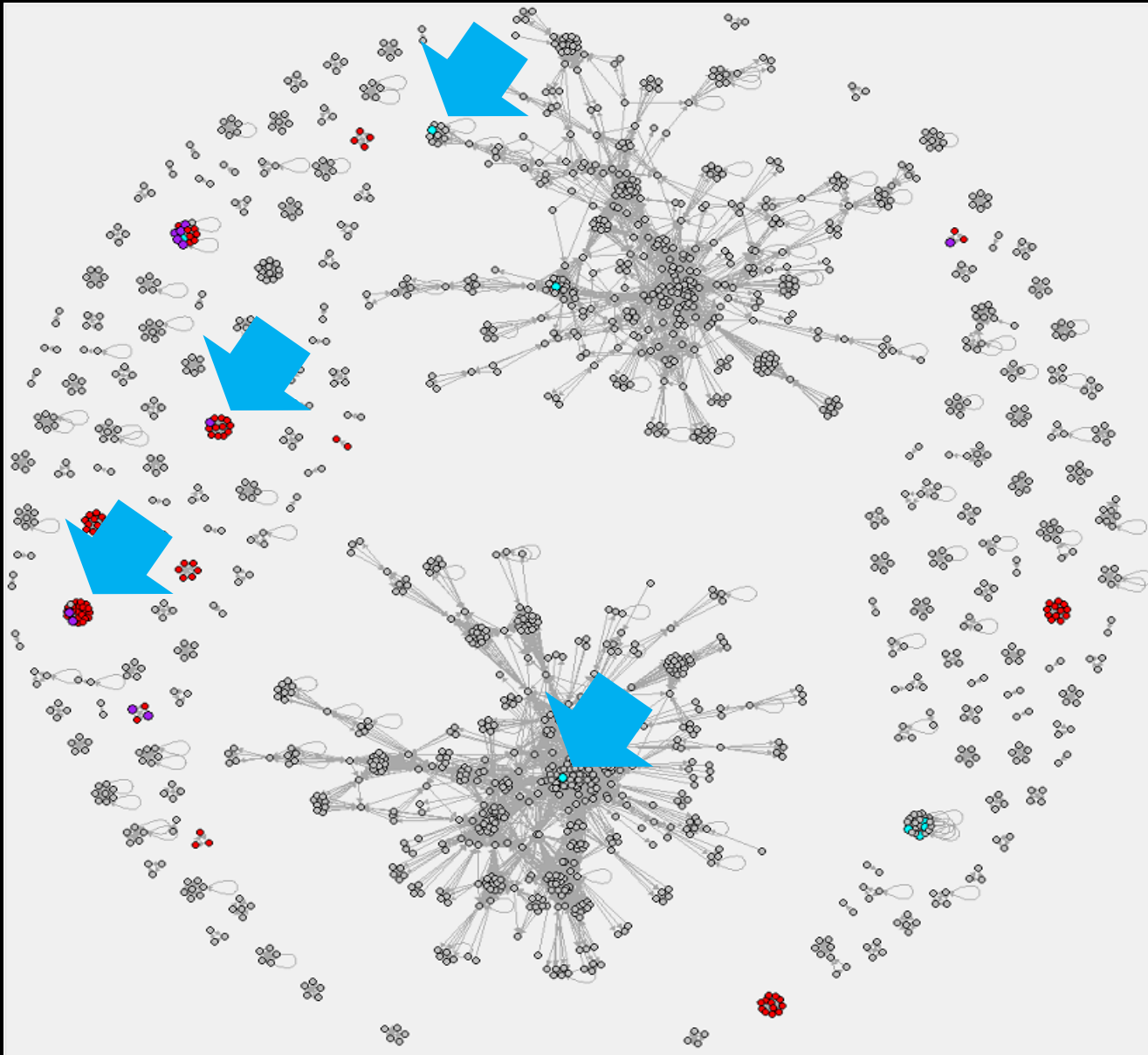
$$B(x) = B_0(x) + \sum_{f \in \text{Friend}(x)} \left(B_0(f) \frac{F(f, x)}{\sum F(f, f')} \right)$$



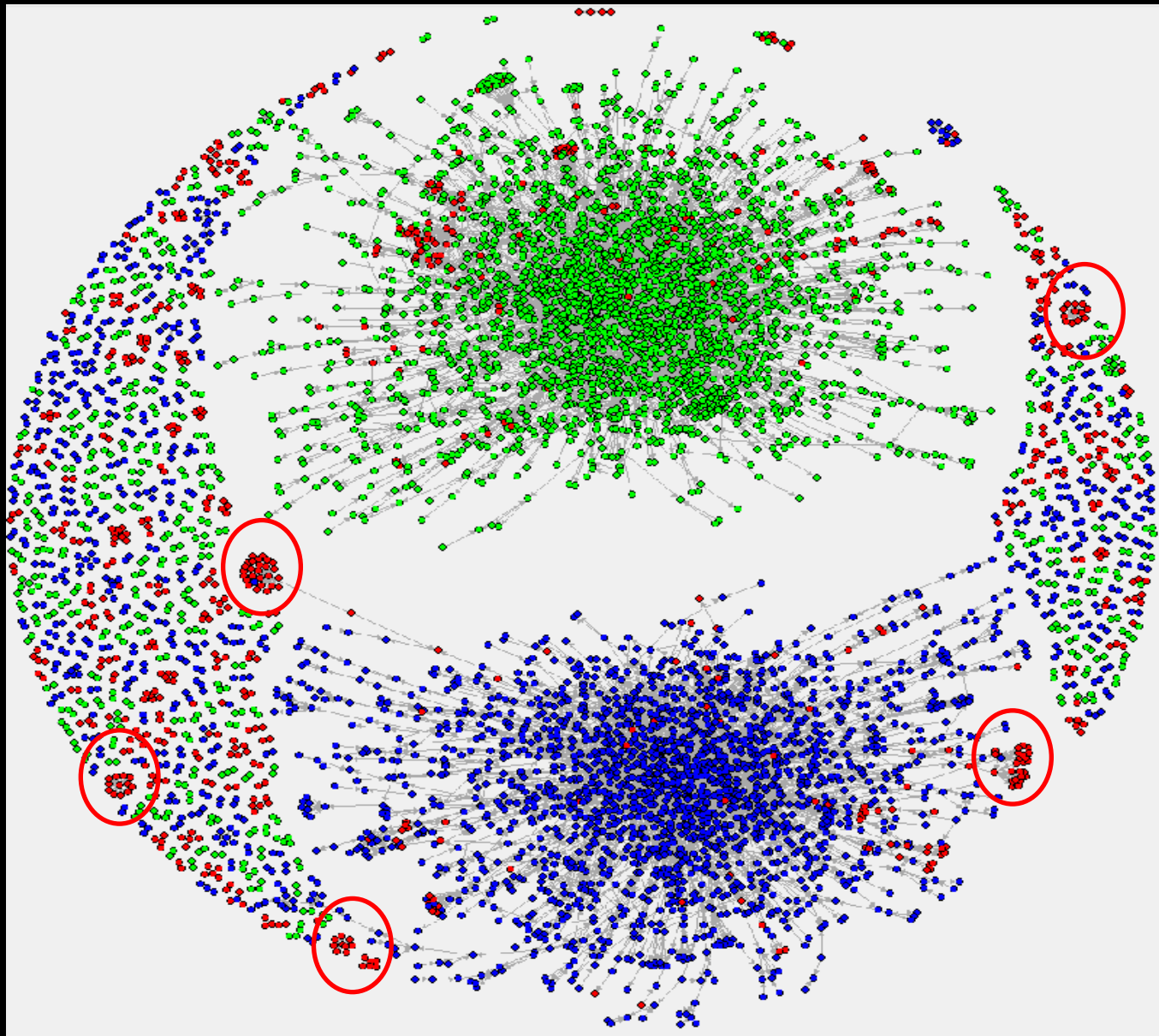
Contagion Model



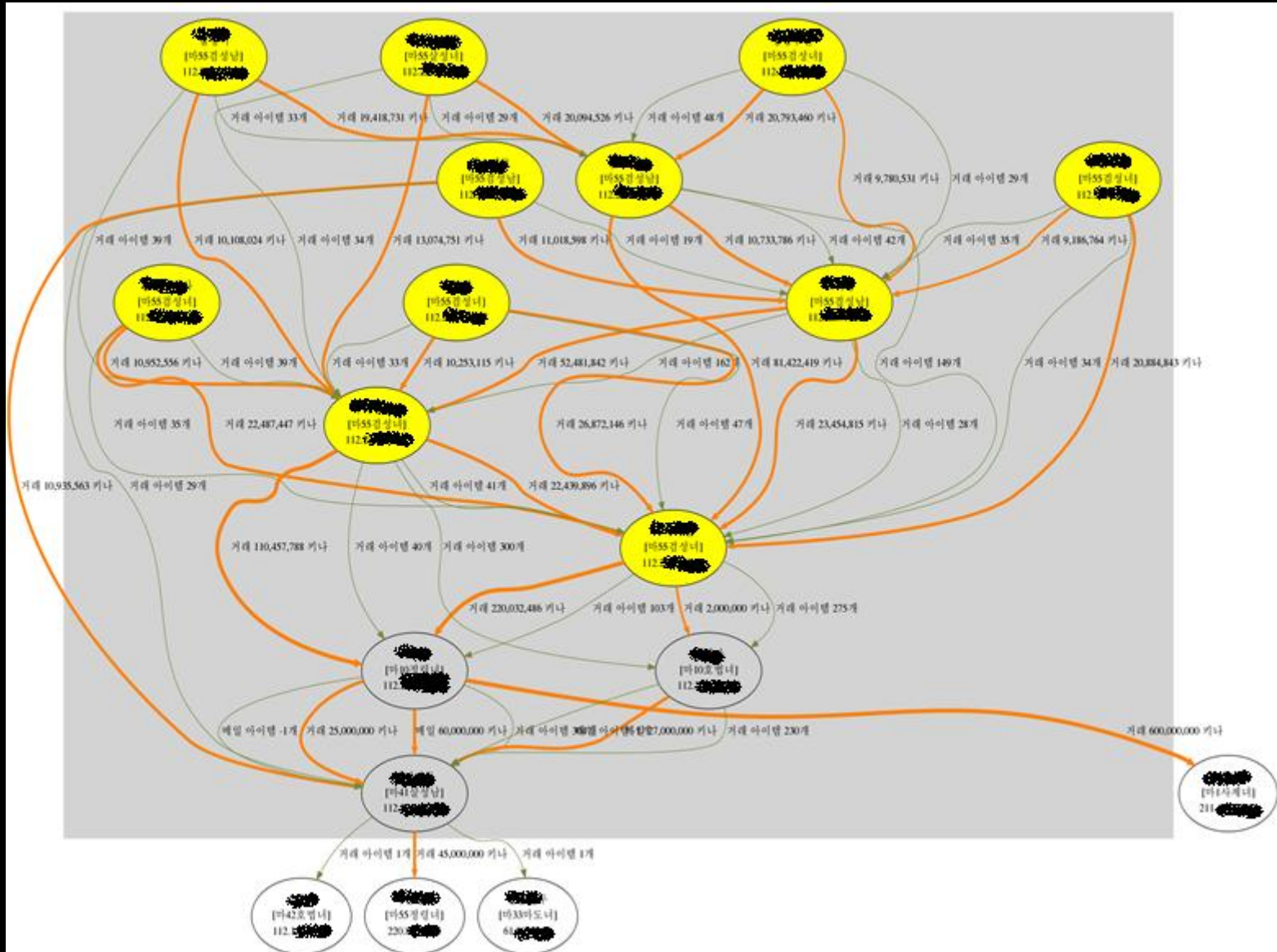
Contagion Model



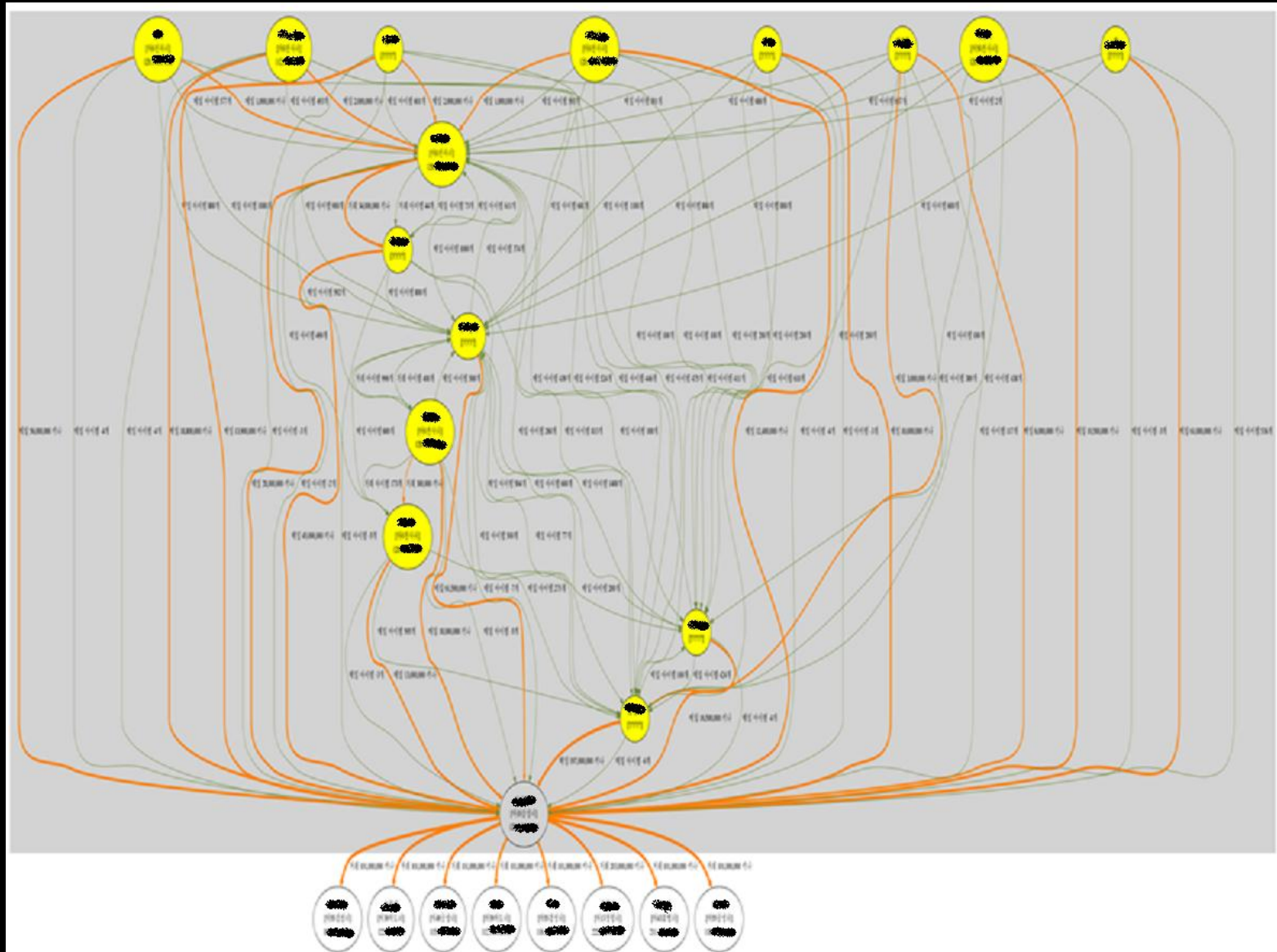
Graph Clustering



Graph Clustering



Graph Clustering



이상 탐지

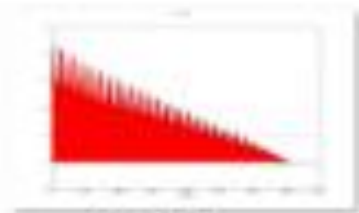
- 게임 버그 활용 플레이 -> 평소와 다른 패턴 발생
 - 아이템 복사 버그
 - 보스 몹 무한 사냥
- 분석 방법
 - 자기 상관 분석
 - 회귀 분석

자기 상관 분석

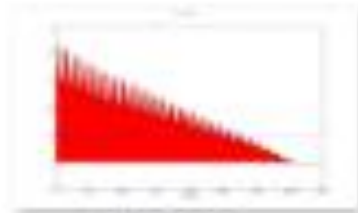
- 로그 횡수 / 정량치 시계열 변동 패턴 학습
- 학습된 변동 패턴에서 크게 벗어나는 시점 탐지

자기 상관 분석

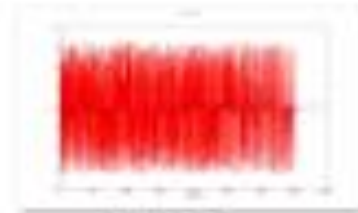
- 자기 상관성 여부 판단



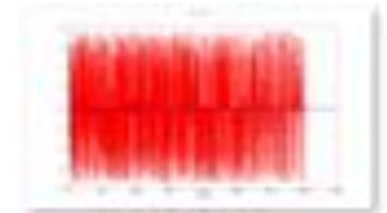
acf.1003.png



acf.1004.png



acf.1005.png



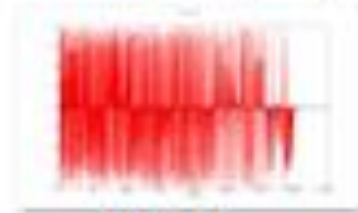
acf.1006.png



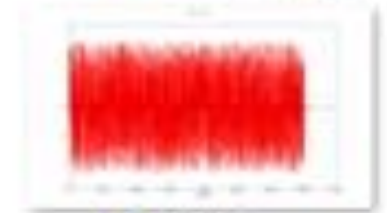
acf.1016.png



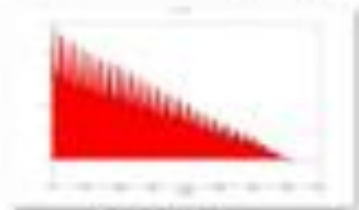
acf.1017.png



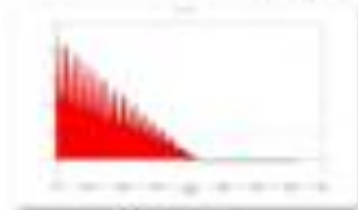
acf.1018.png



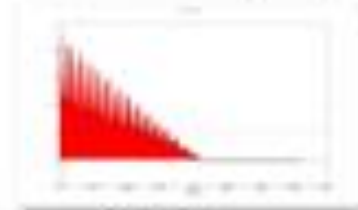
acf.1022.png



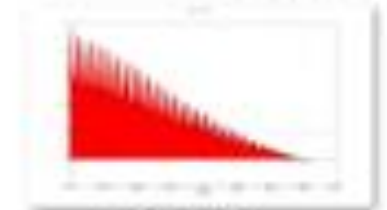
acf.1104.png



acf.1105.png



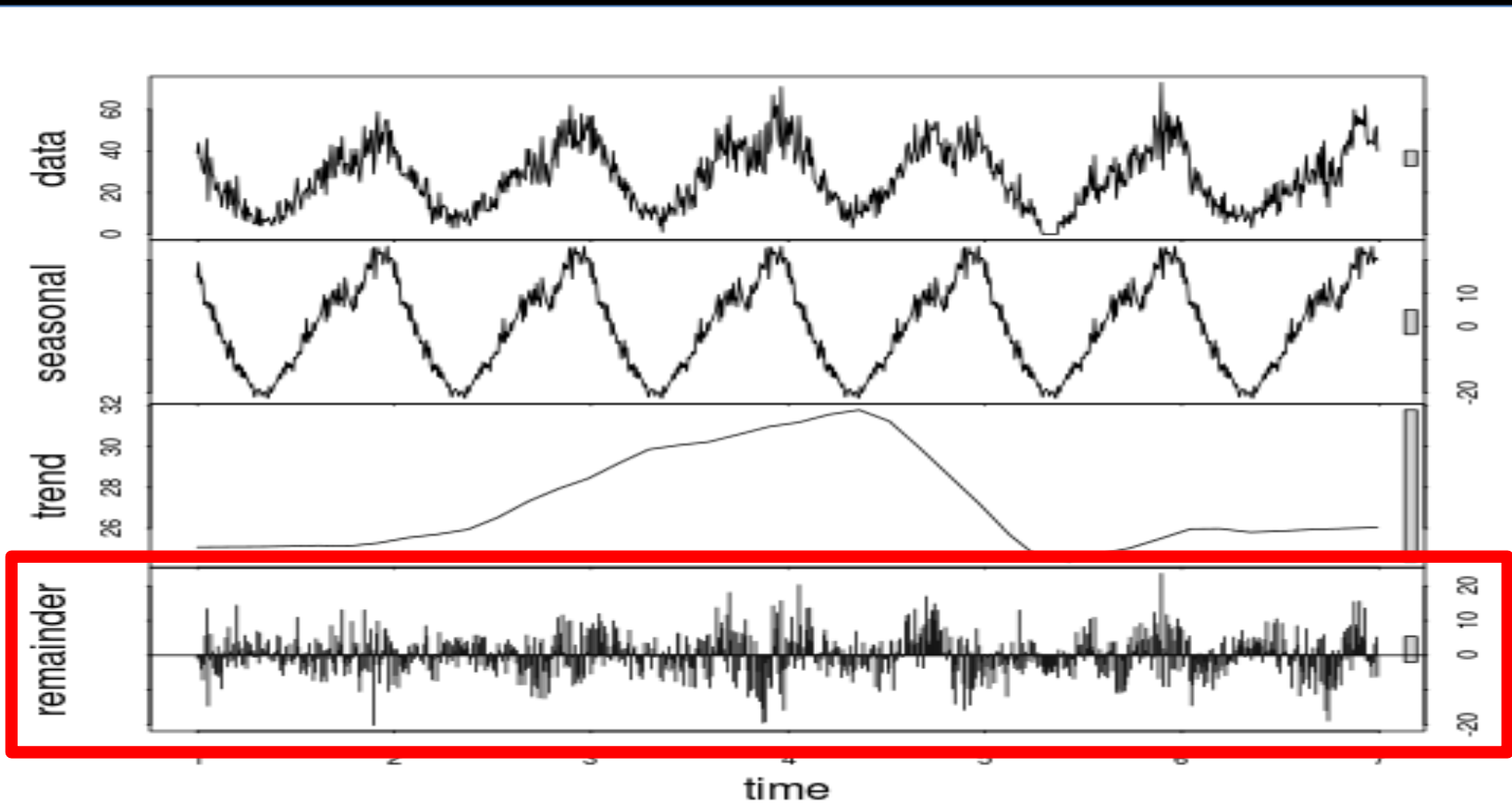
acf.1106.png



acf.1107.png

자기 상관 분석

- **STL** 분석
- Remainder 의 **jitter** 탐지



회귀 분석

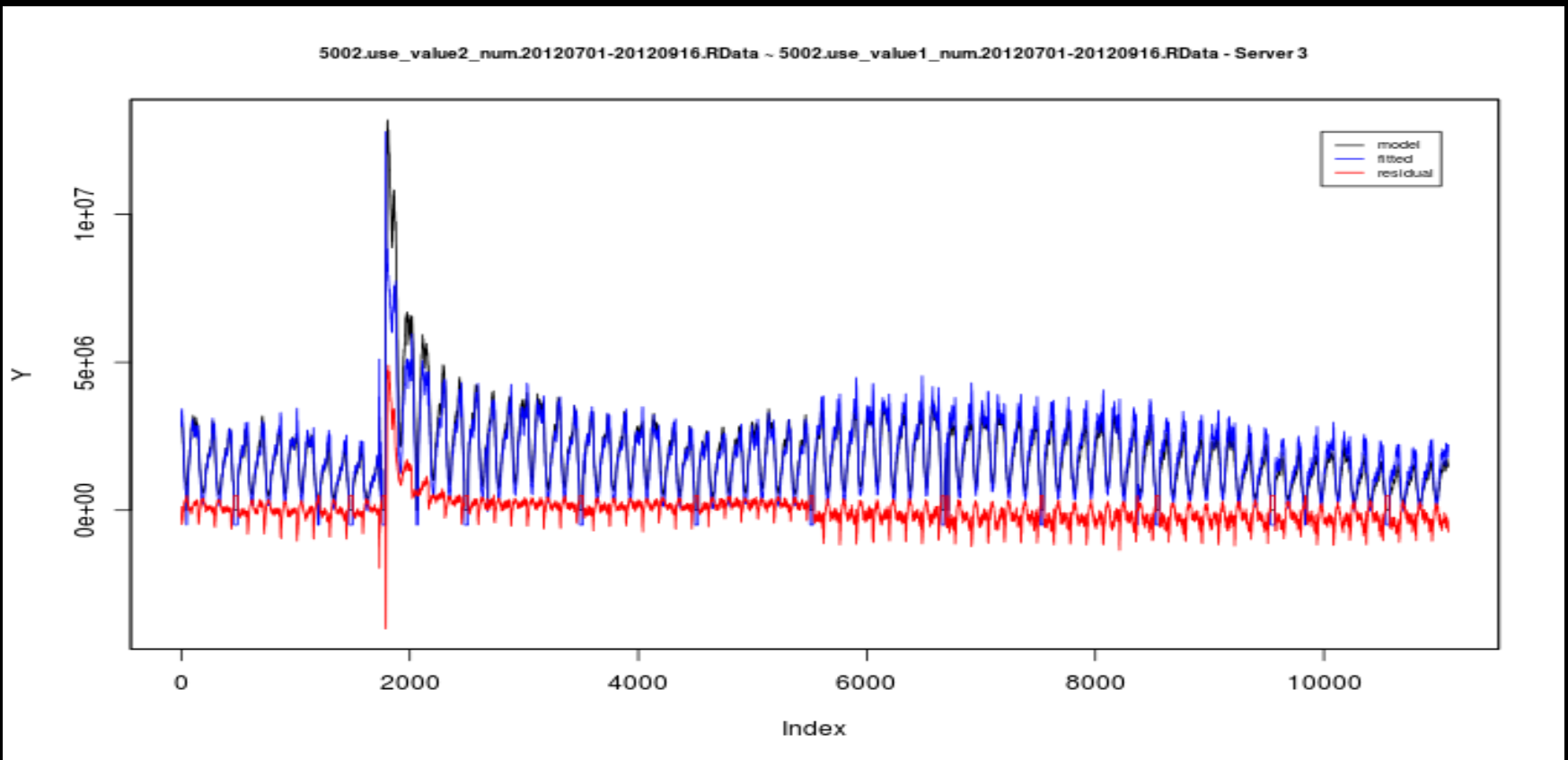
- 같은 시점에 발생하는 경향 강한 로그 쌍 추출
- 로그 쌍에 대한 회귀 모형 생성
- 모형의 예측 값과 실제 값이 크게 어긋나는 시점 탐지

회귀 분석

- 로그 쌍 추출
 - 로그 분포의 왜도와 첨도 자질로 이용
 - Clustering -> Cointegration Test

회귀 분석

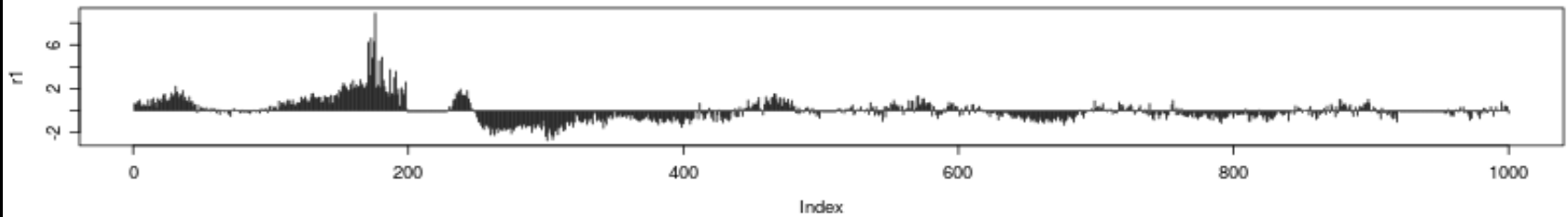
- 회귀 모형 생성
- 예측치와 실측치의 **잔차(residual)** 추출



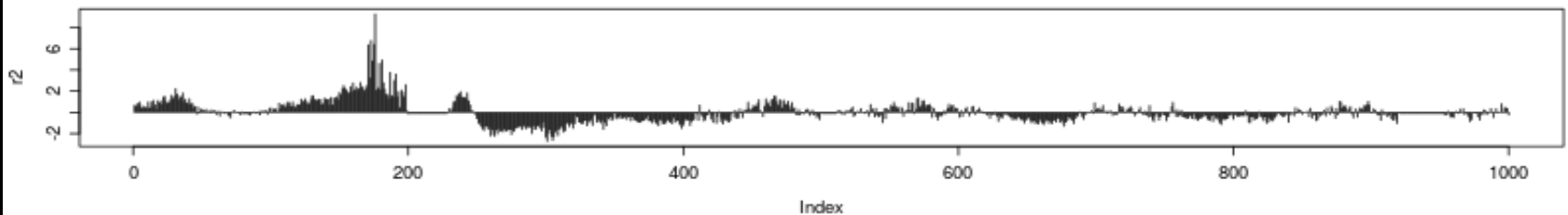
회귀 분석

- 잔차에 대한 jitter 탐지

Standardized residuals



Studentized residuals



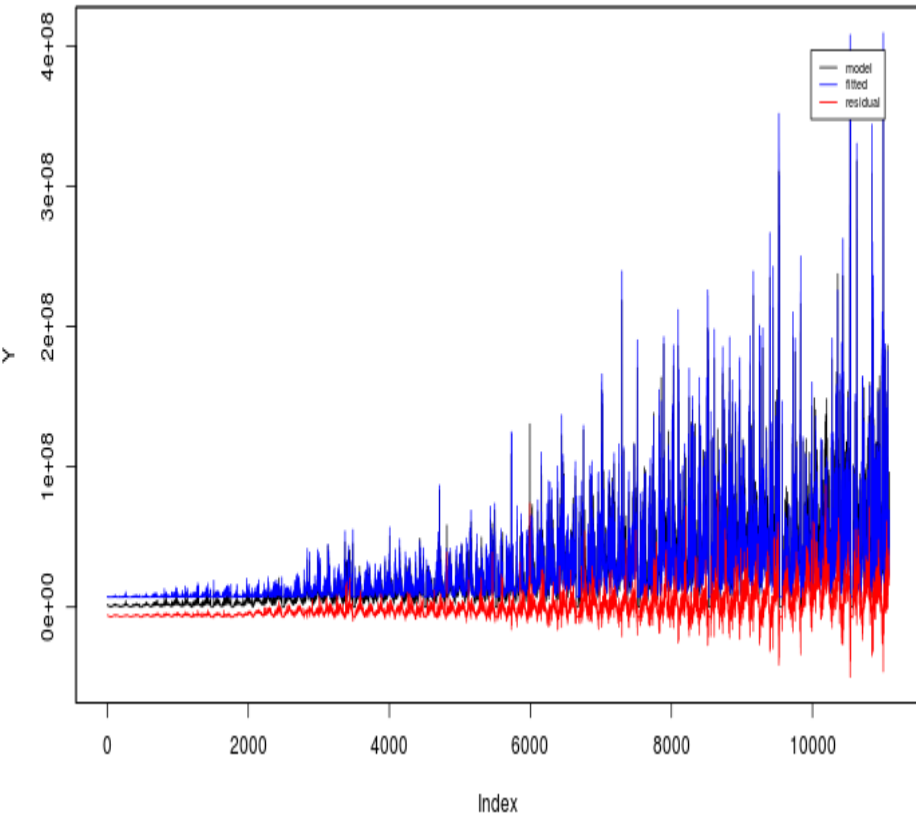
Cook's distance



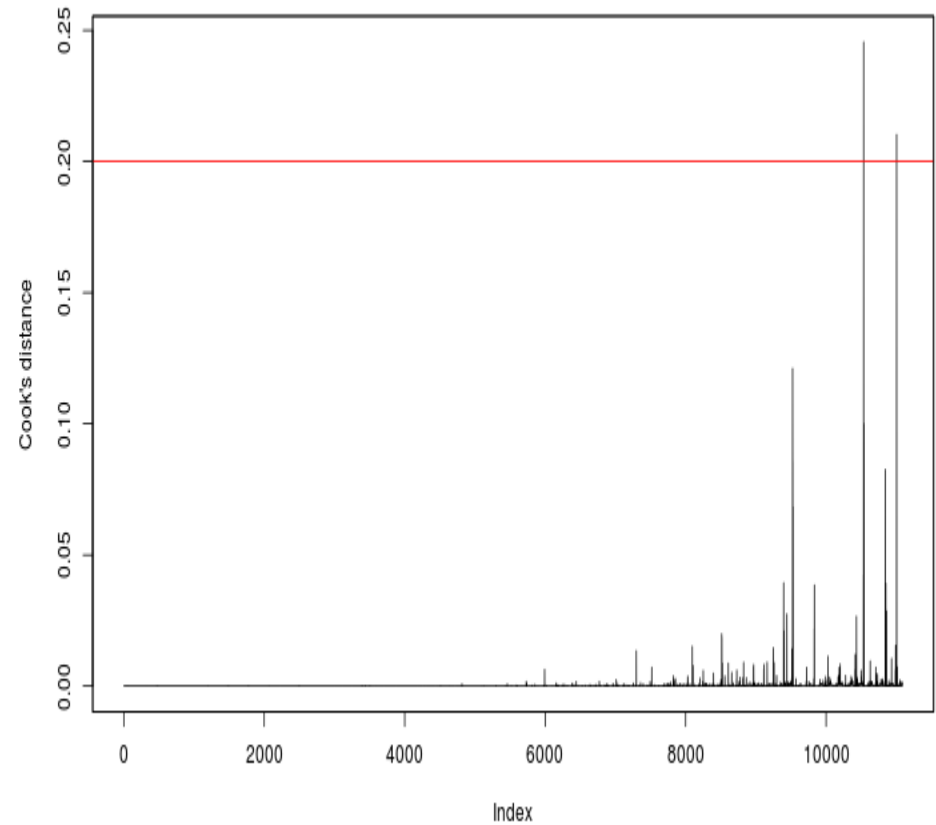
회귀 분석

- 잔차에 대한 jitter 탐지

2206.use_value2_num.20120701-20120916.RData - 1018.use_value2_num.20120701-20120916.RData - Server13



2206.use_value2_num.20120701-20120916.RData - 1018.use_value2_num.20120701-20120916.RData - Cook's distance - Server13



결론

- 게임 데이터는 활용성 및 가능성이 매우 높은 데이터
- 활용성 및 효율성 극대화 -> **실질 가치 창출** 목표
 - 대규모 데이터 처리 인프라 및 플랫폼 구축
 - 심화 분석을 위해 R 도입
- R의 장점
 - 높은 **유연성**
 - 다양한 **패키지**
 - 다른 개발 툴과의 **연동 및 협업** 유리

결론

- 빅 데이터와 R
 - 빅 데이터 분석은 8할이 데이터 정제
 - 정제 후 데이터는 stand alone, in-memory 처리
 - 대규모 데이터 정제 -> Hadoop
 - 정제 데이터 분석 -> R

Q&A