Network exploration

### installing/loading libraries

# install.packages("BiocManager")   
# BiocManager::install("WGCNA")   
  
packs <- c("sna", "tsna", "ndtv", "tidyverse")  
for (i in 1:length(packs)){  
 if(packs[i] %in% rownames(installed.packages()) == FALSE){  
 install.packages(packs[i], repos="http://cran.univ-paris1.fr/")  
 }  
 library(packs[i], character.only = TRUE)  
}

### data

indir <- "/Users/susana.martinez/Documents/LAO\_TS\_analysis/MGE\_analysis/Host\_CRISPR\_MGE\_analysis/Networks"  
  
# phage networks  
nodes\_ph.in <- paste(indir, "Final\_nodes\_spreadMGMT\_phage.tsv", sep = "/")  
net\_ph.in <- paste(indir, "Final\_rbin\_phagenet.tsv", sep = "/")  
  
# plasmid networks  
nodes\_pl.in <- paste(indir, "Final\_nodes\_spreadMGMT\_plasmid.tsv", sep = "/")  
net\_pl.in <- paste(indir, "Final\_rbin\_plasmidnet.tsv", sep = "/")

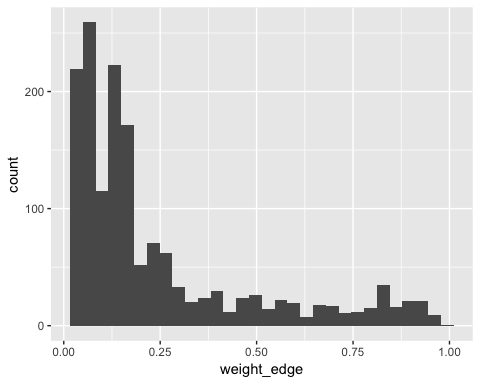
### PH-host network: interactions

* Get edge attributes, i.e. present/absence og edges per time point

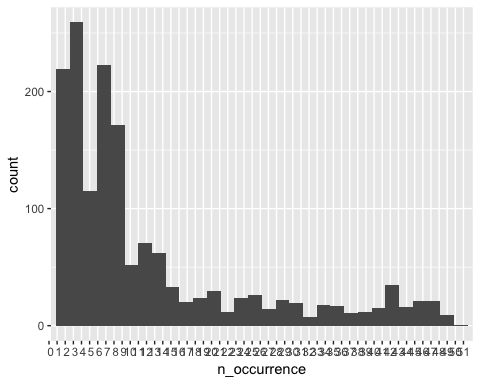
# get unite(A-B)/dates with 0 or 1, and then get average, median, sd  
# which will be then the wight of the edge, variance, etc.  
  
nodes\_ph <- read\_tsv(nodes\_ph.in) %>% gather(date, omic, -node, -type) %>% filter(!is.na(omic))  
net\_ph <- read\_tsv(net\_ph.in)  
  
net\_ph %>%   
 inner\_join(., nodes\_ph, by = c("rbin" = "node")) %>% filter(!is.na(omic)) %>%   
 select(-omic, -type) %>%  
 inner\_join(., nodes\_ph, by = c("pspcc" = "node", "date")) %>% filter(!is.na(omic)) %>%   
 select(-omic, -type) %>%   
 unite("edge", c("rbin", "pspcc"), sep = "-") %>%   
 mutate(interaction = 1) %>%   
 spread(date, interaction, fill = 0) -> edge\_tab  
#write\_tsv(edge\_tab, paste(indir, "ALL\_phage\_networks", "Edge\_attributes.tsv", sep = "/"))  
  
edge\_tab %>%   
 gather(date, presence, -edge) %>%   
 group\_by(edge) %>%   
 mutate(n\_occurrence = sum(presence)) %>%   
 mutate(weight\_edge = mean(presence)) -> edge\_tab\_weigth\_occ  
write\_tsv(edge\_tab\_weigth\_occ, paste(indir, "ALL\_phage\_networks", "Edge\_attributes\_witgh\_dates.tsv", sep = "/"))

* Estimate a weight per edge based on the averaged presence

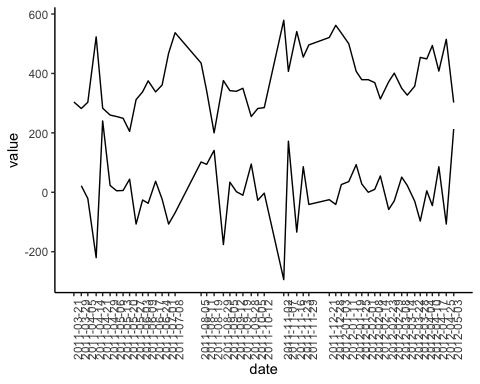
edge\_tab\_weigth\_occ %>%   
 spread(date, presence) %>%   
 select(weight\_edge) %>%   
 ggplot(aes(x=weight\_edge)) +   
 geom\_histogram()



edge\_tab\_weigth\_occ %>%  
 spread(date, presence) %>%   
 select(n\_occurrence) %>%   
 ggplot(aes(x=n\_occurrence)) +   
 geom\_histogram() +  
 scale\_x\_discrete(limits=0:51)

 + Difference in the number of new edges per time point: with respect to the previous time point

TPs <- edge\_tab\_weigth\_occ %>% ungroup() %>% select(date) %>% unique() %>% mutate(date = as.Date(date))  
  
edge\_tab\_weigth\_occ %>%   
 ungroup() %>%   
 mutate(date = as.Date(date)) %>%   
 group\_by(date) %>%   
 mutate(n\_ints = sum(presence)) %>%   
 ungroup() %>%   
 select(date, n\_ints) %>% unique() %>%   
 mutate(t\_minus\_1 = lag(n\_ints)-n\_ints) %>%  
 gather(type, value, -date) %>%   
 group\_by(type) %>%   
 mutate(average=mean(value, na.rm=TRUE), median = median(value, na.rm=TRUE), variance = sd(value, na.rm=TRUE)) -> n\_ints\_with\_lag  
  
n\_ints\_with\_lag %>%   
 ggplot(aes(x=date, y=value, fill=type)) +  
 geom\_line() +   
 scale\_x\_date(breaks=TPs$date) +  
 theme\_classic() +   
 theme(axis.text.x = element\_text(angle = 90))

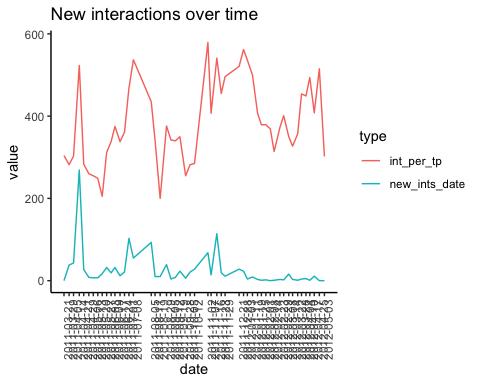


n\_ints\_with\_lag %>% ungroup() %>% select(2,4,5,6) %>% unique()

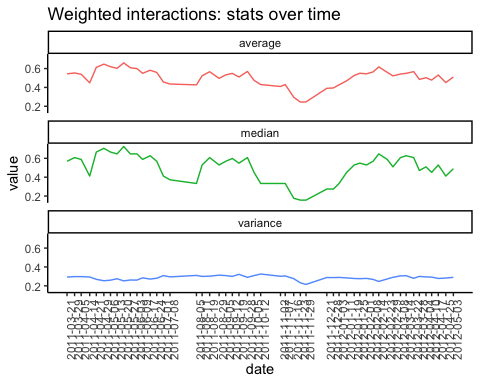
## # A tibble: 2 x 4  
## type average median variance  
## <chr> <dbl> <dbl> <dbl>  
## 1 n\_ints 382. 369 98.0  
## 2 t\_minus\_1 0.04 3.5 97.5

* new edges per time point, i.e. edges that did not appear until TPn –> new interactions over time

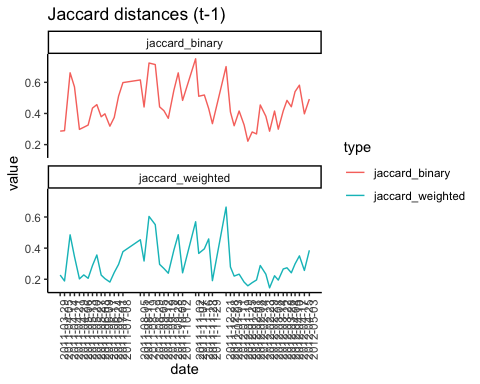
edge\_tab\_weigth\_occ %>%   
 ungroup() %>%   
 select(1,2,3) %>%   
 group\_by(date) %>%   
 mutate(int\_per\_tp = sum(presence)) %>% ungroup() %>%   
 group\_by(edge) %>%   
 mutate(cummulative\_presence\_ints = cumsum(presence)) %>%   
 mutate(new\_ints = case\_when(cummulative\_presence\_ints == 0 ~ 0,  
 cummulative\_presence\_ints == 1 & lag(cummulative\_presence\_ints) == 0 ~ 1,  
 cummulative\_presence\_ints == lag(cummulative\_presence\_ints) ~ 0,  
 cummulative\_presence\_ints > 1 ~ 0)) %>% ungroup() %>%   
 select(edge, date, new\_ints, int\_per\_tp) %>% unique() %>%   
 group\_by(date) %>%   
 mutate(new\_ints\_date = sum(new\_ints, na.rm = T)) %>%   
 ungroup() %>%   
 select(date, new\_ints\_date, int\_per\_tp) %>% unique() %>%   
 gather(type, value, -date) %>%   
 mutate(date = as.Date(date)) %>%   
 ggplot(aes(x=date, y=value, colour=type)) +  
 geom\_line(show.legend = T) +   
 scale\_x\_date(breaks=TPs$date) +  
 theme\_classic() +   
 theme(axis.text.x = element\_text(angle = 90))+  
 ggtitle("New interactions over time")

 + Weighted nteractions per time point: avergae, median, sd, per time point and show over time

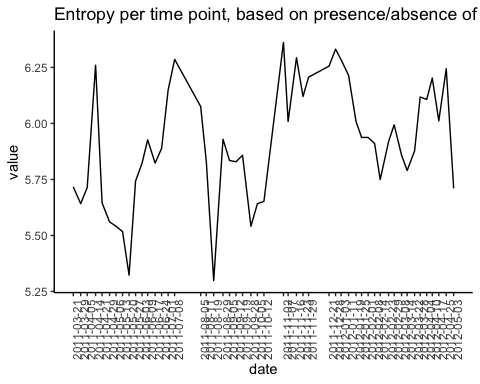
edge\_tab\_weigth\_occ %>%   
 ungroup() %>%   
 mutate(weight\_edge = ifelse(weight\_edge\*presence == 0, NA, weight\_edge\*presence)) %>%   
 select(edge, date, weight\_edge) %>%   
 group\_by(date) %>%   
 mutate(average = mean(weight\_edge, na.rm = T), median = median(weight\_edge, na.rm = T), variance = sd(weight\_edge, na.rm = T)) %>%   
 ungroup() %>%   
 select(-edge, -weight\_edge) %>% unique() %>%   
 gather(variable, value, -date) %>%   
 mutate(date = as.Date(date)) %>%   
 ggplot(aes(x=date, y=value, colour=variable)) +  
 geom\_line(show.legend = F) +   
 scale\_x\_date(breaks=TPs$date) +  
 facet\_wrap(~variable, ncol = 1) +   
 theme\_classic() +   
 theme(axis.text.x = element\_text(angle = 90))+  
 ggtitle("Weighted interactions: stats over time")

 + Jaccard distances between time points, based on the presence/abscence of edges and the weight when there is edge

# get matrix to calculate distances  
# presence/absence  
# weighted network  
edge\_tab\_weigth\_occ %>%   
 mutate(weight\_edge = ifelse(weight\_edge\*presence == 0, 0, weight\_edge\*presence)) %>%   
 select(edge, date, weight\_edge) %>%   
 spread(date, weight\_edge) %>%   
 remove\_rownames() %>%   
 column\_to\_rownames("edge") %>%   
 as.matrix() %>%   
 t() -> mt\_pa  
  
library(vegan)  
rm(list=c("jaccard\_pa\_bi", "lag\_dist\_pa\_b", "jaccard\_pa", "lag\_dist\_pa"))  
  
jaccard\_pa\_bi <- vegdist(mt\_pa, method = "jaccard", diag=F, binary=T) %>% as.matrix()  
lag\_dist\_pa\_bi <- jaccard\_pa\_bi[row(jaccard\_pa\_bi) == col(jaccard\_pa\_bi) + 1]  
lag\_dist\_pa\_bi %>%   
 as.tibble() %>%   
 mutate(date = as.Date(colnames(jaccard\_pa\_bi)[-1])) -> lag\_dist\_pa\_bi  
  
###  
jaccard\_pa <- vegdist(mt\_pa, method = "jaccard", diag=F) %>% as.matrix()  
lag\_dist\_pa <- jaccard\_pa[row(jaccard\_pa) == col(jaccard\_pa) + 1]  
lag\_dist\_pa %>%   
 as.tibble() %>%   
 mutate(date = as.Date(colnames(jaccard\_pa)[-1])) %>%   
 full\_join(lag\_dist\_pa\_bi,., by = "date") %>%   
 rename(jaccard\_binary = value.x, jaccard\_weighted = value.y) %>%   
 gather(type, value, -date) %>%   
 ggplot(aes(x=date, y=value, color=type)) +  
 geom\_line(show.legend = T) +   
 scale\_x\_date(breaks=TPs$date[-1]) +  
 facet\_wrap(~type, ncol = 1) +   
 theme\_classic() +   
 theme(axis.text.x = element\_text(angle = 90))+  
 ggtitle("Jaccard distances (t-1)")

 + Entropy

#install.packages("entropy")  
library(entropy)  
edge\_tab\_weigth\_occ %>%   
 select(date, presence) %>%  
 group\_by(date) %>%   
 mutate(entropy = entropy(presence)) %>%   
 ungroup() %>%   
 select(date, entropy) %>% unique() %>%   
 mutate(date = as.Date(date)) %>%   
 gather(type, value, -date) %>%   
 ggplot(aes(x=date, y=value)) +  
 geom\_line(show.legend = F) +   
 scale\_x\_date(breaks=TPs$date) +  
 theme\_classic() +   
 theme(axis.text.x = element\_text(angle = 90))+  
 ggtitle("Entropy per time point, based on presence/absence of interaction")



### PH-host network: nodes

# get node degree, but per part, since this is a bipartite net  
edge\_tab\_weigth\_occ %>%   
 select(edge, date, presence) %>%   
 separate(edge, into = c("A", "B"), sep = "-")

## # A tibble: 80,529 x 4  
## A B date presence  
## <chr> <chr> <chr> <dbl>  
## 1 A01\_O1.2.4 A01\_C1.1.14\_contig\_12673 2011-03-21 0  
## 2 A01\_O1.2.4 A01\_E1.2.10\_contig\_32472 2011-03-21 0  
## 3 A01\_O1.2.4 A01\_E13\_contig\_43485 2011-03-21 0  
## 4 A01\_O1.2.4 A01\_N\_contig\_18941 2011-03-21 0  
## 5 A01\_O1.2.4 A01\_N\_contig\_30400 2011-03-21 0  
## 6 A01\_O1.2.4 A01\_N\_contig\_39579 2011-03-21 0  
## 7 A01\_O1.2.4 A01\_N\_contig\_70118 2011-03-21 0  
## 8 A01\_O1.2.4 A02\_N\_contig\_82475 2011-03-21 0  
## 9 A01\_O1.2.4 D01\_L1.42\_contig\_218719 2011-03-21 0  
## 10 A01\_O1.2.4 D01\_N\_contig\_228630 2011-03-21 0  
## # ... with 80,519 more rows

### Convert the ReGe-PH net to ReGe-ReGe net connected by common CRISPR interactions